

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

v1 protein - protein search, using sw model

run on: January 30, 2004, 13:12:38 ; Search time 22 Seconds
(without alignments)
2313.635 Million cell updates/sec

title: US-09-769-699-2

erfect score: 6294

sequence: 1 MENTQKTVTPTGVLGYVA.....DELFLSGIPDKIGNITMEM 1203

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 328717 seqs, 42310858 residues

total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/aa/PCTUS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3054	48.5	1452	4	US-09-127-227-2
2	183	2.9	35	2	US-08-607-332B-9
3	183	2.9	35	3	US-09-338-876-9
4	131	2.1	757	3	US-09-413-814-84
5	123	2.0	1074	4	US-09-071-035-358
6	123	2.0	1074	4	US-09-071-035-394
7	118.5	1.9	370	4	US-09-252-991A-16913
8	111	1.8	5215	3	US-07-542-734C-4
9	106.5	1.7	3567	3	US-08-439-009A-4
10	106.5	1.7	3567	3	US-07-731-157A-7
11	106	1.7	774	2	US-08-541-780-7
12	106	1.7	774	2	US-09-107-532A-3855
13	105.5	1.7	1095	4	US-08-633-760-46
14	105	1.7	774	1	US-08-633-760-48
15	105	1.7	774	1	US-08-931-608A-5
16	104.5	1.7	635	4	US-08-019-870-5
17	104.5	1.7	774	1	US-08-019-870-3
18	103	1.6	774	1	US-08-070-165F-6
19	103	1.6	729	1	US-08-885-418-6
20	102	1.6	729	2	US-07-747-901A-3
21	102	1.6	774	1	US-07-935-312-3
22	102	1.6	774	1	US-08-633-760-50
23	102	1.6	774	1	US-08-019-870-1
24	102	1.6	774	1	US-08-019-870-6
25	101.5	1.6	773	1	US-08-019-870-8
26	101.5	1.6	774	1	US-08-019-870-8
27	101.5	1.6	774	1	US-08-019-870-8

28	101.5	1.6	774	1	US-08-019-870-11	Sequence 11, Appl
29	101.5	1.6	774	1	US-08-633-760-52	Sequence 52, Appl
30	101	1.6	995	4	US-09-657-931A-1	Sequence 1, Appl
31	101	1.6	1024	4	US-09-582-737-48	Sequence 48, Appl
32	101	1.6	1466	4	US-09-282-537-20	Sequence 20, Appl
33	101	1.6	7257	3	US-09-335-409-5	Sequence 5, Appl
34	101	1.6	7257	4	US-09-568-102-5	Sequence 5, Appl
35	101	1.6	7257	4	US-09-567-969-5	Sequence 5, Appl
36	101	1.6	7257	4	US-09-568-480-5	Sequence 5, Appl
37	101	1.6	7257	4	US-09-568-486-5	Sequence 5, Appl
38	101	1.6	7257	4	US-09-568-472-5	Sequence 5, Appl
39	101	1.6	7257	4	US-09-567-899-5	Sequence 5, Appl
40	100.5	1.6	1321	1	US-08-261-822A-3	Sequence 3, Appl
41	100.5	1.6	1321	5	PCT-US95-07744A-3	Sequence 3, Appl
42	100.5	1.6	2616	6	5206163-3	Patent No. 5206163
43	99.5	1.6	600	4	US-09-252-991A-29817	Sequence 29817, A
44	99	1.6	915	1	US-08-328-322-5	Sequence 5, Appl
45	98.5	1.6	801	4	US-09-351-150A-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-09-127-227-2
; Sequence 2, Application US/09127227
; Patent No. 6399354
; GENERAL INFORMATION:
; APPLICANT: David M. Knipe
; APPLICANT: Travis J. Taylor
; APPLICANT: Elizabeth McNamee
; TITLE OF INVENTION: Replication-Competent Virus Expressing A
; FILE REFERENCE: HU98-05
; CURRENT APPLICATION NUMBER: US/09/127,227
; CURRENT FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1452
; TYPE: PRT
; ORGANISM: herpesvirus
US-09-127-227-2

Query Match	48.5%	Score	3054;	DB 4;	Length	1452;	
Best Local Similarity	49.8%	Pred. No.	2.7e-304;	Indels	48;	Gaps	17;
Matches	609;	Conservative	200;	Mismatches	366;		
Qy	1	MENTQKTVT---VPTGPLGVY--ACRVEDLDEEISFLAARSTDSALLPLMRNLTV	55				
Db	1	METPKPTATIKVPPGPLGVYARACPSEGIEL--LALLSARSGDADVAVAPLVVGLTVE	58				
Qy	56	KFTSSLIANVSGARTTGLAGAGITLKLTTSHFVPSVVFHGGKHVLPSSAAPNLTRACNA	115				
Db	59	SGFEANVAVVGGRTTGLGTAVSLKLTTPSHYSSVVVFHGGKHLDPSTQAPNLTRLCER	118				
Qy	116	ARERFGSRCCGPPVDGAVETTGAEICTRLGLRPENTILYLVVTFALFKEAVFMCNVLHY	175				
Db	119	ARRHFGSDYTPRGDLKHETTGEALCERLGLDPRALLYLVVTEGKFAVCINNTFLHL	178				
Qy	176	GGDLIVHINHGDIIRIFLPVQLFMPDVRNLVPPFNTHRSIGEGFYVTPFVNTGLCH	235				
Db	179	GGSDKVTIGGAEVHRIRFVYPLQLFMPDFSRVIAEPFNANHSIGENTYLPFPFNRLNR	238				
Qy	236	LHDCVITAPMAVALVRNVTAARGAHAFLAFDENHGAALPPDITTYTFOSSSGTGTAR	299				
Db	239	LLFEAVGPPAAVALRCRNVDAVARAAHAFDENHGAALPADITTFATFASQ--KTPR	299				
Qy	296	GARRNDVNSTKSPSGGFFERRLASIMAADTALHAEIFNTGIYEEPTDITKEWPMFIGM	351				
Db	297	GGR-----DGGKGAGGAGGFEORLASVWAGDAALAESIVMAVDEPPTDITSAWPLCEGQ	351				
Qy	356	EGTLPRLNALGSYTARVAGVICAMVFPNSALYLTEVEDSGWTEAKDGGPGGPFNRFYQF	415				

09/769,699
2.21.04
SEARCH NOTES

352 DTAARANAVGAYLARAAGLVGVFSTNSALHTEVDGADPADPKDHSK-PSFYREFLV 410
416 AGPHLANPQDTRCHVL-----SQSTGSSNTEFSDYDYLALICFGAPILARLLFYL 468
411 PGTHVAANPQDREGHVVGFEGRPTAPLVGGTQ-ETAGEHLAMLCGFSALLAKMLFYL 469
469 ERCDAGFTGGHG-DALXYVTGTSDSEIPSCLEKTRPVCAHTVHRLRQRPRFCQAT 527
470 ERCDGGVIVGEQMDVFRYVADSNTQDVPNCNLCTFDIRHACVHTLRLARHPKPSAA 529
528 RQPIGVFTMNSQVSDCDPLGNVAPYLILRKPGQOTEAATMOTQTYRATLERLFDLEQ 587
530 RGAIGVFTMNSMYSDCDVLGNVAAFSALKR-ADGSETARTIMQETTYRATERVMAELET 588
588 ERLDRGAPCSSEGLSSVIVDHPTRFRLLDRLARIEQTTTFQFMKVLVEVDYKIREGLS 647
589 LOYVDQAVPTANGRLTITITREALHTVNNRVQVQVREVEQLMNLVEGNEKFRDGLG 648
648 EATHSMALTFDYPGAFCEITNFLVKRTHLAVQDLALSOCHCVYGCQVEGRNFRNQFQ 707
649 EANHAMSUTLDPYACPCPLLQLLGRRLNLAVIDLALSOCHGVFAGQSVGEGNFRNQFQ 708
708 PVLRRFRVDLNGGFISTRSTVTLSEG-PVSAENPTLGDAPAGRTFDGDLARVSVEVI 766
709 PVLRRFRVDMFNNGFLSAKTLTVALSEGAALCAPSLTAGQTPAPAESFEGDVARVTLGFP 768
767 RDIRVKNRVFSGNCTNLSEARARLVGLASAYORQEXRVDMLHGLGFLKQPHGLLFP 826
769 KELRVKSVRLFAGASANASEAARVASLQSAQYKPDKRVLDLGLPLGLLKQPHAAIFP 828
827 RQMPNSKSPNQFWTLLORNOPADKLTHEETITIAVKRFTVEEYAAINFILPTCI 886
829 NGKPGSGNQPNQFWTALORNLPARLLSREDIETIAFIKKFLSDYGAINFILAPNV 889
887 GELACFYMANLTKYCDHSXYLINTLGTIIGARRPRDPSSVLEHWIRKDVTSADDIETQA 946
889 SELAMYMANQILRYCDHASTFYINTLTAIAGSRPPSPVQAAAW---SAOGGAGLEAGA 945
947 KALLEKTENLPELWTAFTSTHLVRAANQPMVVLG:ISIKYHGAAGNNRVFAGHWSG 1006
946 RALMDAVIDAHPGAWTSMFASCNLLRPVMAARPMVVLG:ISIKYGMAGNDRVFAQGNWAS 1005
1007 LINGGNVCPFLFTDRTRFLIACPGGTCIPVTGSSGNRETTLSDOVRGIIVSGAMVQ 1066
1006 LMGGKACPELLIFDRTRKFLVACPRAGFVCAASNLGGGAHSSCEQURGIISGGAVA 1065
1067 LAIYATVVRVAVGARQHVAFDDWLSITDDEFIARDLELHDQIIQTLPTWTVGAL--- 1123
1066 SSVFVATVKSJLGPRTQQLQIEDWALLEDEYLSSEEMELTARALERGNGENSTDAALEVA 1125
1124 -EAVKILDEKTAGDGETTNLAFND--SCFSDHTTSNVLNIGSNISSGTVPGKRP 1180
1126 HEAEALVSQLGNAGE-----VFNFGDGC-----DONATPFGGPGAGPAPAGRA 1173
1181 PEDDELFDLSGIPKHNITMEM 1203
1174 FHGDDPFG-EGPDKKGDJLDM 1195

RESULT 2
US-08-807-332B-9
; Sequence 9, Application US/0807332B
; Patent No. 5959074
; GENERAL INFORMATION:
; APPLICANT: Dreyfus, David H.
; APPLICANT: Gelfand, Erwin W.
; TITLE OF INVENTION: PRODUCTS AND PROCESSES FOR REGULATION OF
; TITLE OF INVENTION: GENE RECOMBINATION
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross
; STREET: 1700 Lincoln St., Suite 3500

CITY: Denver
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/807,332B
FILING DATE: 28-FEB-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kovarik, Joseph E.
REGISTRATION NUMBER: 33,005
REFERENCE/DOCKET NUMBER: 2879-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700
TELEFAX: 303/863-0223
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 35 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-807-332B-9
Query Match 2.9%; Score 183; DB 2; Length 35;
Best Local Similarity 100.0%; Pred. No. 1.4e-11;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 545 DPLGNVAPYLILRKPGDQTEAAKATMQTYRATLE 579
DB 1 DPLGNVAPYLILRKPGDQTEAAKATMQTYRATLE 35
RESULT 3
US-09-338-876-9
; Sequence 9, Application US/09338876
; Patent No. 6187584
; GENERAL INFORMATION:
; APPLICANT: Dreyfus, David H.
; APPLICANT: Gelfand, Erwin W.
; TITLE OF INVENTION: PRODUCTS AND PROCESSES FOR REGULATION OF
; TITLE OF INVENTION: GENE RECOMBINATION
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/338,876
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/807,332
FILING DATE: 28-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Kovarik, Joseph E.
REGISTRATION NUMBER: 33,005
REFERENCE/DOCKET NUMBER: 2879-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700

TELEFAX: 303/863-0223

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 35 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

IS-09-338-876-9

Query Match 2.9%; Score 183; DB 3; Length 35;

Best Local Similarity 100.0%; Pred. No. 1.4e-11;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

545 DPLGNVAPYLILRKPGDQTEAAKATMDTYRATLE 579

1 DPLGNVAPYLILRKPGDQTEAAKATMDTYRATLE 35

RESULT 4

JS-09-413-814-84

Sequence 84, Application US/09413814

Patent No. 6225064

GENERAL INFORMATION:

APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH

APPLICANT: Bristol-Myers Squibb, Co.

APPLICANT: Beyer, Stefan

APPLICANT: Bioecker, Helmut

APPLICANT: Brandt, Petra

APPLICANT: Cino, Paul M

APPLICANT: Dougherty, Brian A

APPLICANT: Goldberg, Steven L

APPLICANT: Hufle, Gerhard

APPLICANT: Mueller, Joachim

APPLICANT: Reichenbach, Hans

TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or

TITLE OF INVENTION: heteropolyketide compounds

FILE REFERENCE: PCT/US 99/23535

CURRENT APPLICATION NUMBER: US/09/413,814

EARLIER FILING DATE: 1999-10-07

EARLIER FILING DATE: 1999-10-09

NUMBER OF SEQ ID NOS: 107

SOFTWARE: Patent in Ver. 2.1

SEQ ID NO 84

LENGTH: 757

TYPE: PRT

ORGANISM: Sorangium cellulosum

JS-09-413-814-84

Query Match

Best Local Similarity 21.1%; Pred. No. 0.001;

Matches 146; Conservative 84; Mismatches 251; Indels 212; Gaps 37;

QY

54 VEKFTSSLAWSGART-----TGLAGAGITLKLTTSHFYSVFVFGGKHVLPSSAAPNL 109

129 LERTLPRAVSIVADARAALTSASAVIVAS-----LPASAAA-- 167

DB

110 TRACNAAREFGSCGCPVVDGAVETTGAEICTRLGLEPENTILYLVVTLFKEAVFMC 169

QY

168 -----ALQRLRWATQPS-PGPIEGFGA-----ALRPESV----- 198

DB

170 NVFLHY-----GGLDIVHNGDVIRPLFFVQLFMPDVNRLPDPFNTHRSIGSGFVY 224

QY

199 -AFLOYTSGSTGEPKGVMLTHGNDL-----LHNSRLIAHGFDLTSPDPV-----GVIW 244

DB

225 PTPYNTGLC-----HLIHDCVIAFMAVALR-----VENVTAV-----ARGAHLAFD-- 267

QY

245 LPYHDMGLIGILQALYRIRVALMGFLPQLQPMKWLRAVSGALGVSGGNPFAYDLC 304

DB

268 -----ENEGAVLPDITVTYFQSSSGTGTARGARRNDVNSTSKPSPSGGFFERRLASIVA 323

QY

305 VRKSEBERALD-----LRSNEVAFPTGAEPPVRAOTLDRFARAPAVSGFRRE--AFYP 355

DB

QY 324 ADTALHAEIFNTGIYEBERTPTDIKEWPMFIENEGTLPRLNALGSYTRVAGVIGAMVF-- 381
DB 356 CYGLAEATLIVSGARAEAPV-----LRLAPEEVLGRAVASAAE--GARVFGV 403
QY 382 -----SPNSALYLTEVEDSGMTEAKDGGPGSFNRFYFAGPHLA-----ANPQTD RDGHV 432
DB 404 SGRALDPRA---VAIVDPAG---NELGFG-EIGEW-VSGPSVAVGYWGRPE----- 447
QY 433 LSSQSTGSSNTSFSDYLALICGFGAPLRLARLLFYLERCDAGFTGGHGDALKYVTGT-- 490
DB 448 -----ETEATFGATLAGSAAPR-----YLRTGDLGFLRGGE-----LFVVGSRK 486
QY 491 -----FDSEIFCSICEKH--TRPVC--AHTTVHRLRQM-----PRFGQATRF 529
DB 487 DLILGRNHFPODIEKTVESSHRAVRPCSAFVSVEHEGERLAVVCEVDPRVAADPRE 546
QY 530 PIGVFTGMSQSDCDPLGNYPYLILRKPGDQTEAAKATMDTYRATLERLFDLEQBR 589
DB 547 IVAAREAVTAHQ-----LVAAHVALIAPGALPKTSSGKVRRECRRAFL-----DALGER 598
QY 590 -----LLDRGAPCSSEGLSSVIVDHPTRRLDRLRARIETTTQFMKVLVETRDYKI 642
DB 599 HVAFAPELDDAGFPDD---APPETEEPSGRSLDRLS-----TLARLRLDAGQIDDL 651
QY 643 ---REGI-SEATHSMALTFDPYSGAFCPITNFI 671
DB 652 PISRFGLDSLAAVELQHAFQVTRGRAIPLTSIL 684

RESULT 5

US-09-071-035-358

Sequence 358, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 358:

SEQUENCE CHARACTERISTICS:

LENGTH: 1074 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-071-035-358

Query Match 2.0%; Score 123; DB 4; Length 1074;
 Best Local Similarity 18.3%; Pred. No. 0.013;
 Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

QY 250 RVNVTAVARGAAHLAFDENHEGAVLPDDITVYFQSSSGTTTARGARRNDVNSTKPS 309
 DB 67 RTTSLYAEYNGAKQTVFC-IEPGVISIPIETHGY-----QKNPL 104

QY 310 PSGGFERRLASIM--AADTALHAENVFNTGIYEE-----TPTDIKWPMFI 353
 DB 105 PMSDKAKLVSLWEKAGTDIDTNMVAQMIWEENVGYKLHSIKRLGGASVDIK-----158

QY 354 GMEGTPLRLNAGSVTAR-----VAGVIG-AMVFPNSALYLTEVEDSGMTEAKDGGP 405
 DB 159 SIEGKINK--AIEEYQKPSFHTVTKILGOSTLIDKXNLNLSFDPKVVQNTA-----211

QY 406 GPSFNRFYQFAGPPLAANPQTRDGHVLSQSSTGSSNTFESVDYLALICGFGAPLARLL 465
 DB 212 ---NIDYRVIGNQLVLP-----NSNSKSGTLLKKSAGTGTP-VAYKK 251

QY 466 FYLERCDAGFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQMPRFQ 525
 DB 252 AGLQVWAGALDKPNTYAIAKINVTKGS-LKIKKIDKESGDIVETVPHL-----DFGK 304

QY 526 A-----TRQPIGVFGTWNQSOYSDCDPLGNYA-----PYLILRKPQDTEAAKATM 570
 DB 305 ALPSKDVITDKDGI-----SILDGIPIHGTQVITKESVPDPYMDITPMAATIKAGETI 358

QY 571 QDYATRIERLIDLEQBRLLDRGAPCSSEGLS-----SVIVDHTP---FRRILDTRLA 621
 DB 359 SMTSNMRKQKQILLEKGT-VETGTDLNDNYSLAGNTFAIRKDSPAGEIVQEIITDEKG 417

QY 622 RIEQTTQFMKVLVETRDYKIREGSEATHSMALTFDP-----YSGAFCEITNFLVKRT 675
 DB 418 RAE--TPRELANALELGYVTE--TKSNGFVNTFKTKVELKYANQVALVTSNVKGO 473

QY 676 HLAVVQDIAL-----SQCHCVFYQQV-----EGRNFRNOFQPVLRFRVFL 717
 DB 474 NQEIETGTTLTREDKDTGNSQCKAEFKGAEVLTFTAKDGOAVKWEAFK-----TEL 526

QY 718 FNGGFSIRTSITVLE-CPYSAPNPTLGO-----DAPAGRTDGDGLARVSVB-----764
 DB 527 VKGTASDETVTLADEKQNVAVXHLAINEYFWQETKAPGYTLDETKYFVSIKKVNE 586

QY 765 ----VIRDIRNVRV-----PSGNTNLSEAA-----RARLVGLASAYQROEKR 805
 DB 587 KNAVITRDVTAKEQVIRGDFDFKAGSADGTAETGENDLSFKVSPLEGTXEITGAEDKA 646

QY 806 VMLHGLGFG-----LLQFHGLLPPGM-----PPNSKS-----PNP 838
 DB 647 TTACNEQLGFDGYGFENLPGYDYLLEIEA---PEGFKITPLEIRSTFKENKDDVAKS 703

QY 839 QWFWTLLORNQ-----MPADKLTHEEITIAAVKEFTBEYAANINLNP--PTCIGE 888
 DB 704 EYVFITEGQKQPKWVTPVEKLTNNE-----FVSVNLMLYDLPEKEDSLTS 754

QY 889 LAQFYMANILKYCHRSOVLINTLSITGARRRDPSSVLHWIRKO---VTSADIETQ 945
 DB 755 LATWDGKNKLANLDPTE-LVDKL-----RYNLHEIKEDYVVAQADIVEA- 799

QY 946 AKALLEKTENLPENLTAFST-----HLVRAAMNORPMVVLGISISKVHGA 992
 DB 800 TQAAQEKDEKAPVIAETATLANKEKTGKWLKLTAEQ-----VLKSVLFPYV 853

QY 993 AGNRRVFOAGNWSGLNGKNVCPLETFDTRRFRFIACPRGGFCIPVTGSSGNRETTLS 1052
 DB 854 YENKVAFEAGNE-----PVA-----KDALSLN 875

QY 1053 QVREGIIVSGGAMVQLAIYATVRAVGAR-----AQHMAFDDWLSLTDDEFLARDL-- 1102
 DB 876 QAG-----TVNCTIERHVSIOQKAHLEDGSGQTFTHGVDNDMFDVSVTHVDLD 923

QY 1103 --BELHQIITLETPWTEGALAEVAKILDE-----NTTAGD-----GETPTNLA FNFD 1149

DB 924 SKSEAFETILVALLPDGINKIWSKGLIEHVNDKEFTKTVLAEKVDTGKYPGKTF- 982
 QY 1150 SCEPSHDTTSNV-----LNISGNSISGTVPGKRPPEDE 1185
 DB 983 -TEINYEKDGNGVNGKHEDLKEKSTLTTPKEVPTIPSTPKQPE 1024

RESULT 6
 US-09-071-035-394
 ; Sequence 394, Application US/09071035
 ; Patent No. 6448043
 ; GENERAL INFORMATION:
 ; APPLICANT: Gil H. Choi
 ; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 ; NUMBER OF SEQUENCES: 436
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/071,035
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: A. Anders Brookes
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB369P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 394:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1074 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-071-035-394

Query Match 2.0%; Score 123; DB 4; Length 1074;
 Best Local Similarity 18.3%; Pred. No. 0.013;
 Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

QY 250 RVNVTAVARGAAHLAFDENHEGAVLPDDITVYFQSSSGTTTARGARRNDVNSTKPS 309
 DB 67 RTTSLYAEYNGAKQTVFC-IEPGVISIPIETHGY-----QKNPL 104

QY 310 PSGGFERRLASIM--AADTALHAENVFNTGIYEE-----TPTDIKWPMFI 353
 DB 105 PMSDKAKLVSLWEKAGTDIDTNMVAQMIWEENVGYKLHSIKRLGGASVDIK-----158

QY 354 GMEGTPLRLNAGSVTAR-----VAGVIG-AMVFPNSALYLTEVEDSGMTEAKDGGP 405
 DB 159 SIEGKINK--AIEEYQKPSFHTVTKILGOSTLIDKXNLNLSFDPKVVQNTA-----211

QY 406 GPSFNRFYQFAGPPLAANPQTRDGHVLSQSSTGSSNTFESVDYLALICGFGAPLARLL 465
 DB 212 ---NIDYRVIGNQLVLP-----NSNSKSGTLLKKSAGTGTP-VAYKK 251

QY 466 FYLERCDAGFTGGHGDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQMPRFQ 525


```

b 252 AGLQTMAGALDKPNTYAIAKINVTGKS-LKIKIDKESGDIVETVPHL-----DFGX 304
y 526 A-----TRQPIGVGTWNSQVSCDPLGNYA-----PYLILRKPQDTEAAKATM 570
b 305 ALPSKDVTTDKDGI-----SILDGIPIGKTKVITEKSVPPPYMIDTTPMAATIKAGETI 358
y 571 QDTVRATLIERIFIDLEQERLLDRGAPCSSEGLS-----SVVDHPT---FRRILDTLRA 621
b 359 SMTSKNMRQKGILLKGTG-VETGTDLWMDNYSLAGNTFAIRKDSPAGEIVQEIITTDK 417
y 622 RIBQTTQFMKVLVETRDYKIREGLSEATHSNALTFD-----YSGAFCDITNPLVKRT 675
b 418 RAE--TPKELANALELGTYYVE-----TKSGNFVNFTKTKVLEKIANQTVALTNSVKGQ 473
y 676 HLAUVQDLAL-----SOCHCVFYGOQV-----EGRNFRNQFQVPLRRRFDVL 717
b 474 NQEIITGETTLTKEDKDTGNSGKAEPNGAEVTLFTAKDQGAQVWSEAFK-----TEL 526
y 718 FNGGFISTRSITVTLSE-GPVSAPNPTLQ-----DAPAGRTDGDGLARVSVE-----764
b 527 VKGTASDETVTLADEKNQAVKHLAINEYFWQETKAPEGTTLDETYPVSIKKVONNE 586
y 765 ----VIRDIRVNRVV-----PSGNCNTLSEAA-----EARLVGLASAYQROEKR 805
b 587 KNAVITRDVTAKEQVIRGDFDFKAGSADGTAETGFNDLSEKVSPLBGTXEITGAEDKA 646
y 806 VMLHGAALGF-----LLKQFHLGPPRG-----PNSKS-----PNP 838
b 647 TTACNEQLGFDGKFNPLPYDYLLEIEA-----PEGFQKITPLEIRSTFKENKDDYAKS 703
y 839 ONFWITLLORNO-----MPADKLTHEITIAAVKRPTEBYAAINFILP---PTCIGE 888
b 704 EVVFIITGEGQKQPKVQVTPVEKLTNNE-----FSVSLNELMLYDLPEKEDSLTS 754
y 889 LAQFYMANLILKYDSOYLINTLSIITGARPRDPSSVLHWIKO---VTSAADITQ 945
b 755 LATWKDGNKNTLDFTE-LVDKL-----RYNLHEIKEDWYVVAQIDVEA- 799
y 946 AKALLEKTENLELMTTAPTST-----HLVRAAMNORPMVVLGISISKYHGA 992
b 800 TKAQEKDKAPVVAETATLANEKTGTWILHKTAEQ-----VLDSIVLNFVY 853
y 993 AGNNRVFQAGNSGLNGGRNVCPLFTFDRTRFIACPRGGFICPVTPGSSGNRETTLS 1052
b 854 YENKVAFAEGNE-----PVA-----KXASLNN 875
y 1053 QVRGIIVSGAMVQAIYATVVRVARG-----AQHAFDDWLSLTDDEFLARDL-- 1102
b 876 QAK-----TVNCTIERHVSIOTKAHLDEGSQTFTHGDMDFDDVSVTHDVL 923
y 1103 --BELHDQIIQTLPTWTEGALEAVKILDE-----KTTAGD---GETPTNLAFND 1149
b 924 GSKEAFETILYALLPDGTNKEIKWSKIEHVNDEKFTKVLAEKVDIGKYPEGTKFTF- 982
y 1150 SCEPSHDTTSNV-----LNTSGNISGSTVPLKRPPEDE 1185
b 983 -TEINVEKDNVNGKHNEDEKESQTLTPKEVETIPSTPKQPE 1024

```

```

; RESULT 7
; JS-09-252-991A-16913
; Sequence 16913, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

```

```

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16913
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-16913

Query Match
Best Local Similarity 1.8%; Score 118.5; DB 4; Length 370;
Best Local Similarity 20.2%; Pred. No. 0.0051;
Matches 93; Conservative 46; Mismatches 137; Indels 185; Gaps 19;

QY 208 POPFNTHESIGEGFYVPTFFNTGLCHLIDCVAPMAVALRVENVTAVARGAAHLAFD 267
Db 17 POPCNAYH-----TDEQWAAVLTED 37
QY 268 ENHEGA-VLPPDITYTFQSSSGTTTARGARNVDN--STSKPSPSGGF-----ER 316
Db 38 AAADGAFVAVKTTGVYCRPSSS---ARRPRENVPEFATAEAAEAAAGYRPSRRAAGDR 93
QY 317 RLASINAAATLHAEIFNTGIYESTPTDIKEMPMFIGME-----GTLPLRNA 364
Db 94 RLAAEQRAERVAQACRMETA---ETPPALEALAAARLUGSPHFHRLFKAETGLTPKAYA 150
QY 365 LGSYTAARVAGVICAMVFPNSALYLTVEDESDGTEAKDGGPGPSFNRYQFAGPFLAANP 424
Db 151 SAYRARELRERLQASASVTEAIY-----DSGFN-----SNSRFEVSSQRLGMRP 196
QY 425 QTRDRGHVLSQSSTGSSNTSFVDYVLIICGAPALLARLLFYLERCDAGFTGGHGDA 484
Db 197 RYDRDG-----GAG-----AAIRFAIGCCSLGAI-----220
QY 485 KYVTGTFDSEIPCSCLEKHTRPVCA-----HTTVHRLRQSMR-----EQQATR 528
Db 221 -----LVAQSGRIGICALLGEEPEPLRELODQFPRAQLLGGDADFELVA 266
QY 529 QPIGVGTWNSQVSCDPLGNYAPVLIIRKPGDQTEAAKATMQ-----DTRATLERLF 582
Db 267 QVVG-----FVESPOLGLDLP-LDVRTGTAQFQVQALREIIPPGSTASYAQIAERI- 316
QY 583 IDLEQERLLDRGAP-----CSSEGLSSVVDHPTFRR 614
Db 317 -----GAPRAVRAVAQACANRIAVAI PCHRVVR 346

```

```

; RESULT 8
; US-09-105-537-2
; Sequence 2, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
; US-09-105-537-2

```

```

Query Match
Best Local Similarity 1.8%; Score 111; DB 3; Length 5215;
Best Local Similarity 19.7%; Pred. No. 4.4;
Matches 250; Conservative 109; Mismatches 419; Indels 494; Gaps 66;

QY 51 NLTVKTFSTSLAV-----SGARTGLAGAGTTLKLTSHFYPSPVFPVHGKHLVP 102
Db 166 SLTVDAQSSSLVAVHLACBSLRAGESTTALV-AGVNL-----NILA 206

```

103 SAAAPNUTRACNAARERFGSRGCGPPVDGAVETTGAEICTRIGLEPENTILYLVVTALF 162
 207 ESA-----VTEERFG-----GLSPDGYATFDARA-----231
 163 KEAVFMCNVLHVGGLDIVHNGDVIRIPLFPVOLPMPDVNRLVDPDPNTHRSIGEGF 222
 232 -----NGVRGEG-----GGVVVKEL-----SRALADGGRVH-----GV 261
 223 VYPPFVNTGLCHLIHDCVIAAPVALVRVNTAVARGAAHLAFDENHGAFLPDDITYT 282
 262 IRASAVNDG-----ATPGLTVPSRAAQEKVLRAYRKAALDPS-AVQ 303
 283 YFSSSSGTTTARGARENDVNSTKSPSGGFERELASIMAAADTALHAELVFNITGVEET 342
 304 YVELHGT-----PVGDPIEAALGAVLGSARPADEPL-----LVGSA 342
 343 PTDIKEMPMFIGMGTPLRNLGSLYARVAGVIGAMVFPNSALYITEVEDSGMTEAKD 402
 343 KTNVGHLEGAAGIVGLIKTLALG--RRRIPASLNPTPHPIPL-----DTLGLDVED 394
 403 GGPSPFNRFVQFAGPLAANPOTDRD-----GHVLSQSTGSSNTSFSVD 448
 395 G-----LREPHDRELLAGVSSFGMGNTNAHVLSSEGPAQGGEGFGID 438
 449 YLALICGFGAPLLARLLFYLERCDAGA-----FTGGHGDALK-----YVTGTFDSEI-P 496
 439 BETFV-----DSGAALPVPVTVGRGGEALRAQAARLHEAVEADPELAP 480
 497 CSLCEK--HTFPVCAHTV-----HRLRQMPRFQATQPTGIVFGTWNVSQVDCDPLG 548
 481 AALARSIVTTRTVTTHRSVVLAPDRARLLDGLGALAAGTAPGVVGT-----528
 549 NYAPYLILRKPDQTEAAKATWQDYTRATLERFLDLEQERLLDRGAPCSSEGLSSVVD 608
 529 -----PARG-----RLAVLF-----SCQGAQR/GMGV-ELYAA 555
 609 HPTFRILDLARIEQTTQFMKVLVETRD-----YKIRELSEATHSNALT 656
 556 HPATAPADFAVAELDPLDPLRAELVAAGTDLTDRVTQPALFAVEVALHRLVESGVT 615
 657 PDPVSGACPITNFKVTHLAVQDLALSOCHCVFYQQVGEHNFERNQOPVLRFRFVD 716
 616 PDLLAGH-----SVGEISAHVAGV--LSLRDA-----ARLVAAGRMLQALP-----656
 717 LFNGGFISTRITVTLSCGVSAPNPTLQDAPAGRTFDGLARVSVVEVIRDRKNRVV 776
 657 --EGAM--VAVEASEEVL--PHL-----AGREBELSLAAN-----GPRAVV 694
 777 FSGNCTNLSEARARLVGLASAYORQEKRVDMHMGALFLKQFHGLF-----825
 695 LAG-----ABRAVLDAVELLREQGRRTKRLSVSHA-----FHSPLMBFMLDDFRVV 741
 826 -----PR-----GMPNKSXPON-----FWTLQRNOMPADKLTHEEITITIAAV 866
 742 EELDFQERVDVTVTGLPVT-----QWTDPEYM-----VDQV-RRPVFLDAV 787
 867 KFTBEYAAINFILPP--TCIGELAQFMANILKYCDHESQYINTLSITIGARPRD 924
 788 -RTLESGADTFLELGPDPGVCSMAADSV-----RDQEAATAVSALRKG--RP-E 833
 925 PSSVLHWIRKQVTSADTIETOKALLEXTENLPELWTTAFTSTHLVRAAMNQRBMVVLGI 984
 834 POSLLAALTTVFVRGHDVD-----WTAHGSTGTVRVPL-----PTYAFQR 874
 985 SISKYHGAAGNRRVFAQGNWGLNGKVCPLTFDTRERPIIACPRGGFTCPVTGPSSG 1044
 875 ERHWDGAARTAPLTAGR-SGTGAG-----TGPAAG 905
 1045 -----NRFTTLDQVRGIIVSGGANVQLAIYATVRA-VGARA 1081
 906 VTSGEGEGEAGAGGGRPARHETT--ERVRAHVA-----VLEYDDPTFVELGLTF 957

1082 QHMAFDWLS-----LTDDEFIARDLEELHDOIQIOTLETPWTVGEALAVKILDEKTTA 1135
 958 KELGFDLSMSVELRNALVDDTGLPLPSGLLFDH-----PTP-----BALAA--HGLDLTG 1006
 1136 GGETPTNLAENFOSCEPS--HDTTSNVLMSISNLSGTVPGLEKPPEDDELFDL-----1189
 1007 GSGGTG-----SADGIPFATPADTTAEPIAIIIG--MACRYPOGVTSPEL-LWRLVAEG 1056
 1190 ----SGIPIKHG 1197
 1057 RDAVSGLPDRG 1068
 RESULT 9
 US-07-642-734C-4
 ; Sequence 4, Application US/07642734C
 ; Patent No. 5824513
 ; GENERAL INFORMATION:
 ; APPLICANT: Katz, L
 ; APPLICANT: Donadio, S
 ; APPLICANT: Mcalpine, J B
 ; TITLE OF INVENTION: Recombinant DNA Method for Producing
 ; Erythromycin Analogs
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Edward H. Gorman
 ; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
 ; STREET: Park Rd
 ; CITY: Abbott Park
 ; STATE: IL
 ; COUNTRY: US
 ; ZIP: 60064-3500
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA: US/07/642,734C
 ; APPLICATION NUMBER: US/07/642,734C
 ; FILING DATE: 17-JAN-91
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Dancikers, Andreas M
 ; REGISTRATION NUMBER: 32652
 ; REFERENCE/DOCKET NUMBER: 4952 US-01
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 708-937-9396
 ; TELEFAX: 708-938-2623
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3567 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-07-642-734C-4
 Query Match 1.7%; Score 106.5; DB 2; Length 3567;
 Best Local Similarity 19.9%; Pred. No. 6.2; Indels 465; Gaps 62;
 Matches 237; Conservative 104; Mismatches
 QY 125 CQGPVVDGAVETTGAE--ICTR-----IGLEPENTILYLVVTALFKEAVFMCNVLH 174
 DB 1165 CPQDILLAAVEBAGASAVVCAQDAALREALGDEP-----VTAL-----VH 1205
 QY 175 YGGLDIVHNGDVIRIPLFPVOLPMPDVNRLVDPDPNTHRSIGEGVYPTFVNTGLC 234
 DB 1206 AGTL-----TNFG-----SISEVAPEEF-----ASTIAAKTALL 1234
 QY 235 HLHDCVITAPMAVALRV-----RNVTVARGAAHL-AFDENHEGAVLPDDITYT 282
 DB 1235 AVL-DEVLGDRAVEEVYCVSSVAGIWWGAGMAVAAGSAYLDAEAHR-----1282
 QY 283 YFQSSSGGTTTARGARRNDVNSTKSPSG-----GF--ERRIASMAADTALHA-EVFN 335

```

1283 -----RRGRSCTSVATPMPALPGAVDDGYLRRGLRS-LSADRAMTHERVLA 1330
336 TGIYETPTDIKEWPMFI-EMEGTLPRLNALGSYARVAGVGFNSALYITEVED 394
1331 AGPVSAVADV-DWPLVSEGFATR------TALFAELAGR 1365
395 SGWTEAKDGGPSPFNRFYQFAG------PHLAANPQTDROGHVLSOSTGSSNTE 444
1366 GGOAEAPDS-GTGPBPAQLAGLSPDEOENLELVANAVAE------VLGHSAABINVR 1420
445 FSDYDALICGFGAPLIRLLFYLERCDAGATGGHGDALKYVTGTFDSEIPCSLCEKHT 504
1421 RAPSGLDLSINAWALKRL------SASTG-----LRLPASL----- 1452
505 RPYCAHTTTHRLRQ--RMPRFQATRPICVFGTMSQVSDCDPL-----GNVA 551
1453 --VFDHPTVTALAOHLRLARLVGDQAARVVVGAAD-----ESEPIAIVGICRFPFGIGS 1506
552 PYLILR-----KPGDOTEAA----- 566
1507 PEQLWRVLAEGANLTGTFADRGWDIGRLYHPDPNPGTSYVDKGGFLTDADDPFGPFG 1566
567 -----KATMDTYRATLERFLDLQERLLDRGAPCSSEGLSSVIVDHTPRR 614
1567 ITPREALAMDPOORLMLETAWEAVERAGIDPDALRGTDGTGVFGMNGOSYMOQLLAGEAER 1626
615 I-----LDTLRARIET-----TTQFMKLVETRDYKIREGLSEATHSVAL 655
1627 VDGYOGLNSASVLSGRIATYTFWEGPAUTVDTCSSLVGI--HLAOMALRGCSLAL 1684
656 T-----FDPYSGAFCTPINFVLRKTHLAVVDLALSOCHVFGYQQVQVEGRNFRN----- 704
1685 AGCVTVMSDPYT-----FVDFSTQR-----CLA-SDGRCKAFSARADGFALSGVAAL 1731
705 QPQVLRPRFVDFLNGGFISTRSITVTLSEGVSPAPNPTLQODAPAG-----RTPDGDILAR 760
1732 VLEPLGRAR-----ANG-----HQVLAVRGSAVNOQDASGLAAPNGFSQERVIQALAA 1782
761 VSVVEVIRDIVKRVNVFSGNCTNLSAARLVGLASAYORQEK------VMDLH-- 810
1783 SGVPA-ADVDV--VBAHGTGTGLDPIEAG--ALITVQQRDRPLRLGSKVKTIGHTQ 1836
811 -----GALGFLKOPHGLLFRPMFPNPSKSNPQWFTLLQNOQMPADKLTHEBITTIA 864
1837 AAGAAGVIVKVLAMRHGML-PESLHADELSPHDW-----ESGAVEVLRSEVPWPA 1887
865 AVKRFTEEVAAINFIMPLPTCIGELAQFWMANILKYCHDSQVYLINTLTSITIGARRPD 924
1888 GER---PRRAGVSSFGVSGT-----NAHVIVEEAPAEQ-----EAARTERG 1925
925 PSSVLHWIRKDVTSADITQAKALLEKTENLPEL-----WTTAFTSTHL-VRAAMNQR 977
1926 PLFVLVSGRSEAVAA-----QARALAEHLRDTPELGLTDAAMTLATGRARFVRAV--- 1978
978 PMVVLGIS-SKYHGAAGNRRVFOAGNWSGLNGKNCVPLTFDTRTRFTIACPRGFGICP 1037
1979 -----LQDRAGVCAELDALAECR-----PSADAVAP 2005
1038 VTGPSSGNRETTILSDQVRGIIVSGGAMVQLAIYATVAVRAGARQAHMFDMLSLTDDDEF 1097
2006 VT---SAPRXPVL-----VFPQGAQ-----WVG----- 2026
1098 LARDLBELHQIQTLL-----ETPMTVEGALBAVKILDEKTTAGDG 1138
2027 MARDLLESSEVFAESMRCAEALSPTHDM-----KLLD--VVRG DG 2065

```

RESULT 10
 US-08-439-009A-4
 ; Sequence 4, Application US/08439009A
 ; Patent No. 6004787
 ; GENERAL INFORMATION:

```

; APPLICANT: Donadio, S
; APPLICANT: Katz, L
; APPLICANT: McAlpine, J B
; TITLE OF INVENTION: Method of Directing Biosynthesis of
; TITLE OF INVENTION: Specific Polyketides
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steven F. Weinstein
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/439,009A
; FILING DATE: 11-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Casuto, Dianne
; REGISTRATION NUMBER: 40,943
; REFERENCE/DOCKET NUMBER: 4952.US.D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-938-3137
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3567 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-439-009A-4

Query Match 1.7%; Score 106.5; DB 3; Length 3567;
Best Local Similarity 19.9%; Pred. No. 6.2;
Matches 237; Conservative 104; Mismatches 384; Indels 465; Gaps 62;

QY 125 CQGPVVDGAVETTGAE--ICTR-----LGLPEPTILYLVVTALFKAEVFMCNVPLH 174
DB 1165 CPGDLLAAVEEAGASAVVCAQDAALRALGDEP-----VTAL-----VH 1205
QY 175 YGGLDIVHNGDVIRIPLFPVQLFMPDVNRLVPDPFNTHRSIGSGFYVPTFPYNTGLC 234
DB 1206 AGTL-----TNFG-----SISEVAPEEF-----AETIAAKTALL 1234
QY 235 HLHDCVIAPMAVALRV-----RNVAVARGAAHL-AFDENHEGAVLPDITVT 282
DB 1235 AVL-DEVLGDRAREVREYCYSSVAGIWWGAGMAAAYAGSAYLDALAEHR----- 1282
QY 283 YFQSSSGTITAGARRNDVNSTSKPSPG---GF--ERRLASIMAAATLHA-EVIFN 335
DB 1283 -----AGRSCTSVATPMPALPGAVDDGYLRRGLRS-LSADRAMTHERVLA 1330
QY 336 TGIYETPTDIKEWPMFI-EMEGTLPRLNALGSYARVAGVGFNSALYITEVED 394
DB 1331 AGPVSAVADV-DWPLVSEGFATR------TALFAELAGR 1365
QY 395 SGWTEAKDGGPSPFNRFYQFAG------PHLAANPQTDROGHVLSOSTGSSNTE 444
DB 1366 GGOAEAPDS-GTGPBPAQLAGLSPDEOENLELVANAVAE------VLGHSAABINVR 1420
QY 445 FSDYDALICGFGAPLIRLLFYLERCDAGATGGHGDALKYVTGTFDSEIPCSLCEKHT 504
DB 1421 RAPSGLDLSINAWALKRL------SASTG-----LRLPASL----- 1452
QY 505 RPYCAHTTTHRLRQ--RMPRFQATRPICVFGTMSQVSDCDPL-----GNVA 551
DB 1453 --VFDHPTVTALAOHLRLARLVGDQAARVVVGAAD-----ESEPIAIVGICRFPFGIGS 1506

```

552 PYLIR-----KPGDQTEAA----- 566
1507 PEQLWRVLAEGANLTTGPPADRGWDIGRLYHPDPNPGTSYVDKGGFLTDAAFDPGFFG 1566
567 -----KATMODTYRATLERLIDLEQERLLDRGAPCSSESGLSSVIVDHPTEFR 614
1567 ITPREALAMDQOORLMLETAEWEVERAGIDPDALRGDTGVFGVMGQSYWQLLAGEAER 1626
615 I-----LDTLRARIEQT-----TTFPMKVLVETRDYKIREGLSEATHSMAL 655
1627 VDGQGLGNSAVLSGRITAYTFWEGEPALTVDTCSSSLVGI--HLAMQALRRGECSLAL 1684
656 T-----PDPSGAFCPITTNFLVKRTHLAVVQDIALSQCHQVFGQVQVEGRNFRN----- 704
1695 AGVTVMSDPIT-----FVDFSTGR-----GLA--SDGRCKAFSARADGAFALSEGVAAL 1731
705 QFQVLRFRFVLDLNGFGFISTRSITVTLSEGVSPAPNFTLQGDAPAG-----RTFDGDLAR 760
1732 VLEPLSRAR-----ANG-----HOVLAVLRGSVAVNQDQASGLAAPGPSQBRVIRQALAA 1782
761 VSVEVIRDIRVKNRVVSGNCTNLSEAAARLVGLASAYQOEKR-----VDMHLH-- 810
1783 SGVPA-ADVDV---VEAHGTGTGLDPIEAG--ALIIATYGGDRDRPLRLSGVKTNIHTQ 1836
811 -----GALGELLKQFHGLLPPRGMPNPKSPNPQWFMTLLQRNQMPADKLITHEEITIA 864
1837 AAAGAAGVKKVLMRHWML-PRSLHADELSPHIDM-----ESGAVEVLRREVWPA 1887
865 AVKRETEYAINFNLPTTCIGELAQFYMANLKYCDHSOYLNTLTIITQARRPRD 924
1888 GER-----PRAGVSSFGVSGT-----NAHVIVEEAPAEQ-----EAAATERG 1925
925 PSSVLHWIRKQVTSAADETQAKALLEKTENLPEL-----WTTAFTSTHL-VAAAMNQR 977
1926 PLPFLVLSGRSAVVA-----QARALAEHLRTPELGLTDAWNTLATGRAPDVRAAV-- 1978
978 PMVLGISISKYHGAAGNNRVFQAGNWSGLNGKKNVCLFTFDRTRRFFIACPRGGFICP 1037
1979 -----LGDRAGVCAELDALAEGR-----PSADAVAP 2005
1038 VTGPPSGNRETTLSDOVRGIIVSGGAMQALAIYATVYRAVGARAHQWAFDDMLSLTDEF 1097
2006 VT---SAPRKPVL-----VFPQGAQ-----WVG----- 2026
1098 LARDEELHQDIQTL-----ETPWTVEGALEAVKILDEKTTAGDG 1138
2027 NARDLLESEVFAESMSRCAEALSHTDM-----KLID--VVRGDG 2065

RESULT 11
US-07-731-157A-7
Sequence 7, Application US/07731157A
Patent No. 5457032
GENERAL INFORMATION:
APPLICANT: Quax, Wilhelmus J.
APPLICANT: Misset, Onno
APPLICANT: Van der Laan, Jan M.
APPLICANT: Lenting, Hexman B.M.
TITLE OF INVENTION: Mutated beta-lactam acylase genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOLEY GODWARD CASTRO HUDDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE, 4TH FLOOR
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/731.157A
FILING DATE: 19910509
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 90200962
FILING DATE: 18-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: RAE-VENTER PH.D., BARBARA
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: GERO-027/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-494-7622
TELEFAX: 415-857-0663
TELEX: 380816 COOLEY BA
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 774 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas species
STRAIN: S883
US-07-731-157A-7
Query Match 1.7%; Score 106; DB 1; Length 774;
Best Local Similarity 20.2%; Pred. No. 0.4;
Matches 145; Conservative 79; Mismatches 251; Indels 242; Gaps 37;
QY 258 ARGAHLAFDENHGAVALPDDITYTFQSSSGTTRTARGARRNDVNSTKPSGSGPER 316
DB 183 AANALKRYDGGQDLICGPGVEARLEADLAALRPAVDALLKAMGSDAAGGSGNN 242
QY 317 -----RLAS-----INAAATLHAENVFNTGVEETPTDIKEWPMFIGNVETGLPRNLALGS 367
DB 243 WAPQRTATGRPIIAGDHPHFVEI-----PGMYAQHLCADRFDM-IGL--TVP----- 289
QY 368 YPARVAGVIGAMVFPNSA-----LYLTVEDSGMTE----- 399
DB 290 -----GVPGPFPHAHNGKAVCVTHAFWDIHDLYLEQFAEDGRTARFQNEPEFVAMRRD 343
QY 400 --AKDGGPSPNRFYQAGPHLAANPQTDGRDGHVLSQSSGSSNTSFSDVYALICGFG 457
DB 344 RIARVGGADREFDIVETRHGPIVAGDP--LEGAATLRSQVFAETDLSFDCLTMP--G 398
QY 458 APLARLLFYLRCDCAGAPTG--GH-----GDALKYVTGTFSSEIPCSLCEKHTREVCART 511
DB 399 ASTVAQLY-----DATRNGGLIDHNLVAGDVAGSIGHLVRAVFSRPRENGMLFVPGWS 452
QY 512 TVHRLR-----QRMPRFQATQPIGVFGTMNSQY-----SDCDPLGNAPY---- 553
DB 453 GEHEWKGWIPHEAMPR--VIDPPGGLIVTANNRVVADDDHPDYLCTDCHP-----PYRAE 504
QY 554 -----LILRKGQDTEAAKATMODT-----VRATLERLFI--DLEQERL----- 590
DB 505 RIMERLVASFAFVADDAIAHADTSLSPHVLRLARLEALGIQGSPLPAEELRQTLLANDGR 564
QY 591 LDRGAPCSSEGLSSVIVDHPTRRILDTLRARIBOTTTQFMKVLVETRDYKIREGLSEA- 649
DB 565 MDAGSQAASA-----YNAFRRAL-----TRLVTARSQLEQAI 596
QY 650 THSNALT-----PDPSGAFCPITNFLVKRTHLAVVQ-----DLALSQCHVFGQVQVEGRNF 702
DB 597 AHPFAAVPFGVSPGQGVWAVPT--LLRNDNAGMLKQSWDEALSEALS--ATQNLTRGW 654
QY 703 RNQFQPVLRFRFVLDLNGFGFISTRSITVTLSEGVSPAP-----NPTLQODAP 749
DB 655 GEEHRP-----RFTPLSAQFPAAWALL-----NPVSRPIGGDGTVLANGLVPSAGPEAT 705
QY 750 AGRTFDGDLARVSEVIRDIRVKNRVVFS--GNCTNLSEAAARLVGLASAYQOEKRVDM 808
DB 706 YG-----ALSRYVFDVGNWDN-----SRWVV 726

NY 809 LHGALGFLKQFHGLLFRPGMPPNSKSNPQW-----FWTLQRNQMPADKLTHEEI 860
DB 727 FHGASG-----HPASPHYADQNAPMSDCAMVPMYLSWDRIAAEAVTQSQEL 771

RESULT 12
S-08-541-780-7
Sequence 7, Application US/08541780
Patent No. 5935831
GENERAL INFORMATION:
APPLICANT: Quax, Wilhelmus J.
APPLICANT: Misset, Onno
APPLICANT: Van der Laan, Jan M.
APPLICANT: Lenting, Herman B.M.
TITLE OF INVENTION: Mutated beta-lactam acylase genes
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
STREET: COOLEY GOWARD CASTRO HUDDLESON & TATUM
CITY: FIVE PALO ALTO SQUARE, 4TH FLOOR
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/541,780
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/731,157
FILING DATE:
APPLICATION NUMBER: EP 90200962
FILING DATE: 18-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: RAE-VENTER PH.D., BARBARA
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: GERO-027/00US
TELEPHONE: 415-494-7622
TELEFAX: 415-857-0663
TELEX: 380816 COOLEY PA

INFORMATION FOR SEQ ID NO. 7:
SEQUENCE CHARACTERISTICS:

LENGTH: 774 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas species
STRAIN: SE83

IS-08-541-780-7

Query Match 1.7%; Score 106; DB 2; Length 774;
Best Local Similarity 20.2%; Pred. No. 0.4;
Matches 145; Conservative 79; Mismatches 251; Indels 242; Gaps 37;
258 ARGAAHLAFDENHGAVALPPDITYTFSSSGTTTARGARRNDVNSTKPSGSGFER 316
183 AANALXRYDDGQDILLCPGVEAELEADLAALRPVADALLKAMGGDASDAAGGSGNN 242
317 -----RLAS---INAAFDALHAEVFNFTGIYEETPTDIKEMPFEGTLPRLNALGS 367
243 WAPAPORTATGRPLAGDHRVFEI---PMTAQHHLACDRDM-IGL--TVP----- 289
368 YTARVAGVICAMVFPNSA-----LYLVEVDSGMT----- 399
290 -----GVGFPHFAHNGKVAVCVTHAFWDIHDLYLEQPAEDGRTARFGNEPFPVWRD 343

QY 400 --AKDGGPSENNRYQFAGPHLAANPQDTRDCHVLSQSTGSSNTEFSVDYLALICGFG 457
DB 344 RIAVRGADREFDIVETRHPVIAQDP---LEGAALTURSVOFAETDLSFDCLTNRP--G 398
QY 458 APLLARLLFYLERCDAGAFG--GH-----GDALKVVTGTDFDSEIEPCSLCEKHTRVCAHT 511
DB 399 ASTVAQLY-----DATRQWGLIDHNLVAGDVAGSIGHLVRAVPSRPRENGWLVPGWS 452
QY 512 TVHRLR-----QRMPPFGQATQPIGVFGTWNISQY-----SDCDPLGNYPAY--- 553
DB 453 GEHEMRGWIPEAMPR---VIDPPGGLIVTANNRVVADDDHPDYLCCTDCHP-----PYRAE 504
QY 554 ----LILRKPGQTEAAKATMODT-----YRATLERLFI--DLEQERL----- 590
DB 505 RIMERLVASPAFVADDAAHADTSLSPHVGLRLARLEALGIQGSIPAELELRTLIANDGR 564
QY 591 LDRGAPCSSEGLSSVIVDHTFRRIDTLRARIEQTTTFQMKVLVETRDYKIREGLSEA- 649
DB 565 MDAGSQAASA-----YNAFFRAL-----TRLVTARSLEQAI 596
QY 650 THSMALT---PDPSGAPCPITNPLVKRTHLAVQ-----DLALSQCHCVFYQQVQVGRNF 702
DB 597 AHPFAAVPPGVSPQGVWAVPT--LLRNDGAGMLKGNWDEALSEALS--ATQNTLGRGW 654
QY 703 RNQFQVLRRRFVDFLNFNGGFISTRSITVTLSEGPVSAP-----NFTLGQDAP 749
DB 655 GEEHRP-----RTHPLSAQFPAAWALL-----NPVSRPIGGDGTVLANGLVPSAGPEAT 705
QY 750 AGRTFDGLARVSVEVIRIVKNRVVPS--GNCTNLSEARARLVGLASAYQRQKRVDM 808
DB 706 YG-----ALSRYVDFDVGWMDN-----SRWVY 726
QY 809 LHGALGFLKQFHGLLFRPGMPPNSKSNPQW-----FWTLQRNQMPADKLTHEEI 860
DB 727 FHGASG-----HPASPHYADQNAPMSDCAMVPMYLSWDRIAAEAVTQSQEL 771

RESULT 13

US-09-107-532A-3855
Sequence 3855, Application US/09107532A
Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: PC

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneka

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277
 INFORMATION FOR SEQ ID NO: 3855:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1095 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: Enterococcus faecium
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (B) LOCATION 1..1095
 SEQUENCE DESCRIPTION: SEQ ID NO: 3855:
 US-09-107-532A-3855

Query Match 1.7%; Score 105.5; DB 4; Length 1095;
 Best Local Similarity 18.9%; Pred. No. 0.86;
 Matches 145; Conservative 112; Mismatches 298; Indels 211; Gaps 35;

551 APVILLRKPGDTEAAKATQDYYRATLERFLDLEQRLLDGAPCSSEGLS-----S 604
 360 APTIDTPMTTITKAGETIVTSKNAQKQILDKSG-VETGSLMNDNYSLAGNTFA 418
 605 VIVDHPFRRLD-TLRARIEQTTPQMKVLVETRDYKIREGLSEATHSWALTFDP--- 659
 419 IRKDSPTGEIVQEMTTDENGHAETPREINALELGIYYVE--TKASHGFVNTFKVKE 476
 660 --YSGAFCPTNLFVRLKTHLVVQDIAL-----SQCHCVFYQQV-----EGR- 700
 477 LKVAQNVAVLTVSNVQKQNEVTEGTTLTREDKDTGDKAKGVAFEGTGYTLFTAKDGRA 536
 701 -NFRNQPOVLRFRFVDFLFGNGISIRSIIVTLSEGVSAF-----NPTLGDD--APAGR 752
 537 VKSEAFKP-----EMVGTGKASBETVTLADENKQRAVKHLAINVEYVWGETRAPEGY 589
 753 TFGDGLARVSE-----VIRDVKNRVV-----FSGNCTNLSBA----- 788
 590 TLDETKVPVSIKKVDNEKNNAVITRDVTAKEQIRFGDFPFKPAAGTAETGENDLTF 649
 789 -RRLVGLASAYQKQKRVMDLHGAIGF-----LLKQFHGILLPRGM--- 829
 650 KVSPLGNETITGAEDTATYNEQGLGDFYGFKNFLPYGDYLLLEEVA---PEGFQKIT 706
 830 PPNKSKSP-----NPQFWTLLQRNQ-----MPADKLTHEEITIAAVKRFTEE 872
 707 PLEIRSTFKENKEDFVKSEIVFVITEQDQKPIKTVTVPEKLTN-----KAFSVS 757
 873 YAINFINLP--PTCIGELAQFYMANILKYCDHSQVLYINTLTSITGARRPRDPSSVLH 930
 758 LNRLMLYDLPEEEDSLTSLATWKDGNKELTSLDSTE-LVDKL-----SYNLH 803
 931 WIRKD--VTSAADETQAKALLEKTENLPETLTFTASTHLVRAAMN-----QRPVW-- 980
 804 EIKEDVWVQAQIDVDA-TKAAQEKDEKAPV-VIAETSLANKETGKWTQKHLTAE 861
 981 -VIGISISKYHGAAGNNRVFQAQNGSLNGKKNVCPLEFTRTRFIIACPRGGFICPVY 1039
 862 QVLNKTIVLFNVYENKEAFEAG-----KQVA 889
 1040 GPSSGNRETTLSQVRIIYSGGMVQALAIYATVAVRAGRAQMAFDLWLSLTDDEFLA 1099
 890 -----KDVSLNNAQVTSCTVEHHVSIQTKAHLENG-----SQTFTGVDVDFDVSIT 939
 1100 RDL-----EELHQDIQIOTLETPWTVEGALEAVKILDE-----KTT-----AGGETPTNL 1144
 940 HDVLGSGKEAFETILYALLPDGINKETWKSGLFYEVDNKEFTKTVLAKVDYDGYKPEGT 999
 1145 AFNDSCEPSHDTTSN-----VLNISGNSIGSTVPLKRPPEDE 1185
 1000 KFTFAEINYDKDGTINGKHEDLKEKSQTLTPKEVPTILTSPKQPE 1045

RESULT 14
 US-08-633-760-46
 ; Sequence 46, Application US/08633760
 ; Patent No. 5804429
 ; GENERAL INFORMATION:
 ; APPLICANT: NIWA, MINEO
 ; APPLICANT: SAITO, YOSHIMASA
 ; APPLICANT: FUJIMURA, TAKAO
 ; APPLICANT: ISHII, YOSHINORI
 ; APPLICANT: NOGUCHI, YUJI
 ; TITLE OF INVENTION: A NEW CEPHALOSPORIN C ACYLASE
 ; NUMBER OF SEQUENCES: 64
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OBLON, SPIVAK, MCQUELLAND, MAIER & NEUSTADT,
 ; ADDRESSEE: P.C.
 ; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE 400
 ; CITY: ARLINGTON
 ; STATE: VIRGINIA
 ; COUNTRY: USA
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/633,760
 ; FILING DATE: 01-MAY-1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: OBLON, NORMAN F.
 ; REGISTRATION NUMBER: 24,618
 ; REFERENCE/DOCKET NUMBER: 18-929-0 PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 413-3000
 ; TELEFAX: (703) 413-2220
 ; TELEX: 248855 OPAT UR
 ; INFORMATION FOR SEQ ID NO: 46:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 774 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-633-760-46

Query Match 1.7%; Score 105; DB 1; Length 774;
 Best Local Similarity 20.7%; Pred. No. 0.5;
 Matches 146; Conservative 76; Mismatches 241; Indels 242; Gaps 37;

QY 109 LTR--ACNAARERPGFRCQGPVVDGAVETTGAE-ICTR-----LGLEPENTILYLVVTAL 161
 DB 70 LTRRKALGRAAEMLG---ABAAEADILVRLGMEKVCRRDFEALGVEAKD-----M 117
 QY 162 FKEAVFNCVFLHVGGDLIVHINHGDVIRIPLFPVOLFMFDPVNLVDPDPTNTH-----R 216
 DB 118 LRAVAGVNAFLASGA-----PL-PVEYGLLGAE---PEPWEHSAIVMR 159
 QY 217 SIGEGFYVPTFYNTGLCHLHDCVIAAPVAALVRNVTA-----ARGAAHLAPDENHEG 272
 DB 160 RLG-----LLMGAVWFKLWRMLALPVVGAANALKRLRYDDGGRD 197
 QY 273 AV-LPPDITVTFYFSSSGCTTARGARRNDVNSTKSPSGGFER-----RLAS---IM 322
 DB 198 LCLPPGAEDRLAEADLALRPAVDALLKAMGGDASDAAGGNNNNAVAGRTATGPIL 257
 QY 323 AADTALHAEIFNTGIYEETPTDIKEWPMFIEGEGTLPRNLALGYSYARVAGVIGAMVFS 382
 DB 258 AGDHRVFEI---PGMYAQHLLACDREFM-IGL--TVP-----GVPGPPHFA 298
 QY 383 PNSA-----LYLTEVEDSGMTE-----AKDGGPPSPFNRF 412
 DB 299 HNGKVAICVTHAFMDIHDLYLEQFAGEGRTARFGNDPEPVAWSRDRIAVRGADREFDIV 358

```

QY 413 YQAPAGHLAANPOTDRDGHVLSQSTGSSNTSPSVLYLALICGFGAPLLARLLFVLERCD 472
Db 359 ETHGGPVIAGDP---RDGAALTLSRVQFAETDLSFCLTTP---GASTVAQLY-----D 407
QY 473 AGAFTG--GH-----GDALKVVTGTFDSEIPCSICEKHTRPVCAHTTVHRLR-----QRM 521
Db 408 ATRGWLGLDHNLVAGDVAGSIGHLVRAVPSRPRENGWLPVPGWSGEHWRGWIPIHEAMP 467
QY 522 RFGQATROPICGVGTWNSQY-----SDCDPLGNVAPY-----LILRKPQDQTE 564
Db 468 R---VIDPFGGIIVTANNRVVADHDHDLCTDCHP-----PYRAERIMKRLVANPAFAVD 519
QY 565 AAKATMQDTVRATLERLFIIDLEQERLIDRGA--PCSEGLSSVIV-----D 608
Db 520 DAAAIHADTLSP-----HVGLRRRLEALGARDSDAAGLRQMLVANDGRMDAASEVASA 574
QY 609 HPTFRILDTL---RARIETTTQ-----FMKVLVETRDYKIREG--- 645
Db 575 YNAFRRALTRLVTRDSGLEQAISHPFAAVPGVSPQGVWMAVPTLLRDDDAGMLKGWSW 634
QY 646 ---LSEA-----THSMALTDPVSGAFCP 666
Db 635 DQALSEALSVASQNLTCRSWGEHRRPFRTHPLATQPPAWAGLINP 679

RESULT 15
JS-08-633-760-48
; Sequence 48, Application US/08633760
; Patent No. 5804429
; GENERAL INFORMATION:
; APPLICANT: NIWA, MINBO
; APPLICANT: SAITO, YOSHIMASA
; APPLICANT: FUJIMURA, TAKAO
; APPLICANT: ISHII, YOSHINORI
; APPLICANT: NOGUCHI, YUJI
; TITLE OF INVENTION: A NEW CEPHALOSPORIN C ACYLASE
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESS: P.C.
; STREET: 1755 JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/633,760
; FILING DATE: 01-MAY-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 18-929-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 774 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; JS-08-633-760-48

```

Query Match 1.7%; Score 105; DB 1; Length 774;
 Best Local Similarity 20.7%; Pred. No. 0.5;
 Matches 146; Conservative 76; Mismatches 241; Indels 242; Gaps 37;

```

QY 109 LTR--ACNAAREPGRSROGPPVGVAVTTGAB-ICTR-----LGLEPENTILYLVVTL 161
Db 70 LTRKALGRAAEMLG---AEEAEDILVRLGMEKVCRRDFEALGVEAKD-----M 117
QY 162 PKEAVFNCNVPLHYGGGLDIVHINHGDVIRIPLFPVOLFWPDVNNRLVPDPNTHH-----R 216
Db 118 LRAIVAGVNAFLASGA-----PL-PVYGLLGAE---PEWEPWHSIAVNR 159
QY 217 SIGEGFYPTFPYNTGLCHLIDCVIAPMAVALRVNRVTAV-----ARGAAHLAFDENHEG 272
Db 160 RLQ-----LIMGSVWFKLWMLALPVVGAANALKLYDDGGRD 197
QY 273 AV-LPPDITTYTQSSSSGTTTARGARRNDVNSTSPSPSGGPER-----RLAS---IM 322
Db 198 LLCPPGAEDRLEADLTLRPAVDALLKAWGDSDAAGGSSNNWAVAPGRATGRPIL 257
QY 323 AADTALAEVIFNTGIYEETPTDIKEWPMFIMEGLPLRLNALGYSYTVARVIGAMVFS 382
Db 258 AGDHRVFEI---PGIYAQHHLACDRFDM-IGL--TVP-----GVPGFPHFA 298
QY 383 PNSA-----LYLVEVDSGMT-----AKDGGPGSPNRF 412
Db 299 HNGKVAYCVTHAFMDIHDLYLEQFAGEGRTAREGNDFFPVANSRDRIVRGGADREFDIV 358
QY 413 YQAPAGHLAANPOTDRDGHVLSQSTGSSNTSPSVLYLALICGFGAPLLARLLFVLERCD 472
Db 359 KTRHGPVIAGDP---RDGAALTLSRVQFAETDLSFCLTTP---GASTVAQLY-----D 407
QY 473 AGAFTG--GH-----GDALKVVTGTFDSEIPCSICEKHTRPVCAHTTVHRLR-----QRM 521
Db 408 ATRGWLGLDHNLVAGDVAGSIGHLVRAVPSRPRENGWLPVPGWSGEHWRGWIPIHEAMP 467
QY 522 RFGQATROPICGVGTWNSQY-----SDCDPLGNVAPY-----LILRKPQDQTE 564
Db 468 R---VIDPFGGIIVTANNRVVADHDHDLCTDCHP-----PYRAERIMKRLVANPAFAVD 519
QY 565 AAKATMQDTVRATLERLFIIDLEQERLIDRGA--PCSEGLSSVIV-----D 608
Db 520 DAAAIHADTLSP-----HVGLRRRLEALGARDSDAAGLRQMLVANDGRMDAASEVASA 574
QY 609 HPTFRILDTL---RARIETTTQ-----FMKVLVETRDYKIREG--- 645
Db 575 YNAFRRALTRLVTRDSGLEQAISHPFAAVPGVSPQGVWMAVPTLLRDDDAGMLKGWSW 634
QY 646 ---LSEA-----THSMALTDPVSGAFCP 666
Db 635 DQALSEALSVASQNLTCRSWGEHRRPFRTHPLATQPPAWAGLINP 679

```

Search completed: January 30, 2004, 13:18:32
 Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: January 30, 2004, 13:10:12 ; Search time 46 Seconds

(without alignments)
6748.637 Million cell updates/sec

Title: US-09-769-699-2

Perfect score: 6294

Sequence: 1 MENTQKTVTPGVLGYVA.....DELFDLSGIPKIGNITWEM 1203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.23.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_prodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4627	73.5	1194	12 Q9E1Y7	Q9E1Y7 cercopithec
2	3583	56.9	1208	12 Q39273	Q39273 equine herp
3	3371.5	53.6	1203	12 Q89549	Q89549 bovine herp
4	3327.5	52.9	1177	12 Q82611	Q82611 pseudorabie
5	3012	47.9	1197	12 Q89101	Q89101 herpes simp
6	2796.5	44.4	1190	12 Q9P252	Q9P252 turkey herp
7	2793.5	44.4	1190	12 Q9QTB9	Q9QTB9 marek's dis
8	2788	44.3	1191	12 Q9E6P0	Q9E6P0 turkey herp
9	2781.5	44.2	1190	12 Q9DPO8	Q9DPO8 measles h
10	2764.5	43.9	1190	12 Q9E1G3	Q9E1G3 measles h
11	2710	43.1	1191	12 Q9IBU6	Q9IBU6 turkey herp
12	1445	23.0	999	12 Q9QH63	Q9QH63 gallid herp
13	683.5	10.9	1132	12 Q40913	Q40913 kaposi's sa
14	675	10.7	1133	12 P88904	P88904 kaposi's sa
15	652.5	10.4	1103	12 Q99226	Q99226 murid herpe
16	651.5	10.4	1103	12 Q41928	Q41928 murid herpe

17	651.5	10.4	1127	12 Q993K9	Q993K9 callitrichi
18	645	10.2	1125	12 Q8B414	Q8B414 porcine lym
19	644	10.2	1125	12 Q8B408	Q8B408 porcine lym
20	633	10.1	1128	12 Q9YTQ7	Q9YTQ7 ateline her
21	623.5	9.9	1133	12 Q8UZD2	Q8UZD2 cercopithec
22	620	9.9	1132	12 Q9WRU1	Q9WRU1 macaca mula
23	618	9.8	1145	12 Q66611	Q66611 equine herp
24	596	9.5	1127	12 Q33360	Q33360 alcelaphine
25	589.5	9.4	1126	12 Q8JYD5	Q8JYD5 porcine lym
26	576.5	9.2	1142	12 Q8B424	Q8B424 porcine lym
27	576.5	9.2	1142	12 Q8B3X5	Q8B3X5 porcine lym
28	561.5	8.9	1134	12 Q99D22	Q99D22 bovine herp
29	505	8.0	1223	12 Q8OS31	Q8OS31 chimpanzee
30	484	7.7	1131	12 Q56282	Q56282 human herpe
31	401	6.4	1281	12 Q85425	Q85425 rat cytomeg
32	237.5	3.8	483	12 Q82172	Q82172 leporid her
33	235	3.7	84	12 Q56864	Q56864 feline herp
34	156.5	2.5	1841	2 Q9FE25	Q9FE25 streptomyce
35	143.5	2.3	178	12 Q64908	Q64908 alcelaphine
36	138	2.2	1869	2 Q8L334	Q8L334 aphanizomen
37	137	2.2	1504	2 Q9ZGA6	Q9ZGA6 streptomyce
38	129.5	2.1	919	5 Q9VA02	Q9VA02 drosophila
39	128.5	2.0	919	5 Q8GV6	Q8GV6 drosophila
40	123.5	2.0	1287	16 Q8EV67	Q8EV67 mycoplasma
41	120.5	1.9	788	16 Q8U9K3	Q8U9K3 agrobacteri
42	120.5	1.9	968	4 Q13025	Q13025 homo sapien
43	120.5	1.9	1555	5 Q8TIQ4	Q8TIQ4 trypanosoma
44	120	1.9	2152	16 Q9EX54	Q9EX54 streptomyce
45	119.5	1.9	701	16 Q92EA7	Q92EA7 listeria in

ALIGNMENTS

RESULT 1

Q9E1Y7 ID Q9E1Y7 PRELIMINARY; PRT; 1194 AA.

AC Q9E1Y7
01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SBDNA binding protein.

OS Cercopithecine herpesvirus 7.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=35245;

EN [1]

RP SEQUENCE FROM N.A.

RA Gray W.L., Starnes H.B., White M.W., Ashburn C.V., Mahalingam R.;

RT "Complete Sequence of the Simian Varicella Virus Genome."

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF275348; AAG27202.1; -

DR InterPro; IPR000635; Viral_DNA_bind.

DR Pfam; PF00747; viral_DNA_bd.1

SQ SEQUENCE 1194 AA; 131968 MW; EBA7F3C841965897 CRC64;

Query Match 73.5%; Score 4627; DB 12; Length 1194;

Best Local Similarity 71.8%; Pred. No. 0;

Matches 867; Conservative 143; Mismatches 178; Indels 20; Gaps 8;

Qy 1 MENTQKTVTPGVLGYVACRVEDLDLEIEFLAARSTDSDLLPLMRNLTVKFTTS 60

Db 1 MEATQKTTTPTGVLGYVACNKEDLDLEISLSARSIDSDLLPLIRNLTVKSFTS 60

Qy 61 SLAVVSGARTTGLAGAGITLKTTSHTFYPFVFGGKRVLPSSAAPNLTRACNAAREFF 120

Db 61 SLAVVSGRTTKTGLAGAGITLKTTSHTFYPFVFGGKRVLPSSAAPNLTRACNAAREFF 120

Qy 121 GPERCQPPVDGAVETGCAICTRLGLEPENTILYLVLTALKEAVFMCVFLHYGGLDI 180

Db 121 GFNMKRGSPVDGAIETGADICNRIHLNPEHAFPLYLIVTSLFKEAVYMCNCSFLYYGGIDN 180

Qy 181 VHNHGVIRIPLFPFVQLEFMPDVNRLVPDPFNTHRSIGEGFYVPTFPFYNTGLCHLIHDC 240

09/769,699 22-04
Search Notes

181 VRIGRAVTRIPFLPIHVFDPVNVRLVDPFNATQSIGENFYPTPLFNGINCNLLYDC 240
241 VIAPMAVALRVNVTAVARGAAHLAFDENHEGAVLPDITVTYFQSSSGTTTARGAREN 300
241 VISPIAVSLIRNVVAVARGAAHLAFDENHEGAVLPDITVTYFQSSSGTNGSNVPRUPRON 300
301 DVNSTKSPSGGPERELASIMAAADTALHAENVTFNTGTYEETPTDIKEWPMFIMGEGTLP 360
301 ----TSKSSQT-GPERELASIMAAADTALHAENVTFNTGTYEETPTDIKEWPMFIMGEGTLP 360
361 RLNALGSYTVARVAVIGAMVFSNSALYLTEVEDSGMTAKOGPGPSNRFVQFAGPHL 420
356 RLNLGSYTVARVAVIGAMVFSNSALYLTEVEDSGMTAKOGPGPSNRFVQFAGPHL 415
421 AANPQTDREDGHVLSQSSGTSNTEFFSYDYLALICGFGAPILARLLFYLERCDAGFTGGH 480
416 AANPQVDRDGCWFVSGTNGSVNTEFNVLDLALVCGSPQLLALLFYLERCDAGFTGGH 475
481 GDALKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRPFGQATROPIGVFGTMNSQ 540
476 SDALQYVMFNSDIPCSLCEKHTRPVCAHTTVHRLRQRPFGQATROPIGVFGTMNSQ 535
541 YSCDCPLNGTAPYLLKPGDQTEAAKATMODTYRATLERLFDLQERLLDRGAPCSSE 600
536 YSCDCPLNGTAPYLLKPGDQTEAAKATMODTYRATLERLFDLQERLLDRGAPCSSE 595
601 GLSSVVDHPTFRILDLTARIEQTTQPMKVLVETRDYKIREGLSEATHSWALFPDYP 660
596 GLNSTIIDHATFRALFVTRVQIQTEQIKVLVETRDYKIREGLSEATHSWALFPDYP 655
661 SGAFCPITNELVXKTHLAVQDLALSCHQVYQGVQVEGNFENQPVLRFRFVLENG 720
656 SGSTCPITYFLVXKTHLAVQDLALSCHQVYQGVQVEGNFENQPVLRFRFVLENG 715
721 GFSTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVFSGN 780
716 GFSTRSITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRDIRVKNRVFSGN 774
781 CTNLSEARARLVGLASAYORQKRVDMHGLGFLLLKQFHGLLFRGMPNPSKSNPQW 840
775 CTNLSEARARLVGLASAYORQKRVDMHGLGFLLLKQFHGLLFRGMPNPSKSNPQW 834
841 FWTLQONQPADKLTREHETIAAVKRFTEVAAINFNLPTCTIGELAQFVMAWLLK 900
835 FWTLQONQPADKLTREHETIAAVKRFTEVAAINFNLPTCTIGELAQFVMAWLLK 894
901 YCHSQYLINTLSITGARRPRDPSVHLWIKOVTSAADITQAKALLEKTENL-PEL 959
895 YCHSQYFINTLTAVITGSKRPNPSSALHWIDKEITPTDLETHARLLIRNTEQMSPTM 954
960 WTTAFTSTHLVRAAMNCRPMVLGISTSKYHGAAGNNRVFQAGNWSGLNGKNVCPLETF 1019
955 WISFMTNLRVAVWQRPVILGISTSKYHGAAGNNRVFQAGNWSGLNGKNVCPLETF 1014
1020 DTRRRFIACPRGGFICVTPGSSGNRETTLSQVRGIIVSGGAWQLAIYATVRAVGA 1079
1015 DTRRRFIACPRGGFICVTPGSSGNRETTLSQVRGIIVSGGAWQLAIYATVRAVGA 1074
1080 RAQMAFDDWLSLTDDEFLARDLIELHQIIOLETPWTEGALVAVKIL-DEKTTAGDG 1138
1075 RAQMAFDDWLSLTDDEFLARDLIELHQIIOLETPWTEGALVAVKIL-DEKTTAGDG 1134
1139 ETPNTLAFNFCSPSHDTTNNVLNIGSNISSGTVP--GLKRPDEDELELSGIPK 1195
1135 SGCGNIAFNFTCNDS-----NTEENAFSPYOLCSGVKRPDDDLLFDWSSLPQK 1186
1196 HGNITMFW 1203
1187 M-SLTMDM 1193

RESULT 2

O39273
ID O39273 PRELIMINARY; PRT: 1208 AA.
AC O39273;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Counterpart of HSV-1 gene UL29 and VZV gene 29.
GN 31.
OS Equine herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10331;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NS80567;
RX MEDLINE=98264497; PubMed=9603335;
RA Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-4.";
RL J. Gen. Virol. 79:1197-1203 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NS80567;
RX Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF030027; AAC59547.1;
DR InterPro: IPR000635; Viral_DNA_bind.
DR Pfam: PF00747; viral_DNA_DP; 1.
SQ SEQUENCE 1208 AA; 130607 MW; 17699FED9238C4CB CRC64;
Query Match 56.9%; Score 3583; DB 12; Length 1208;
Best Local Similarity 56.4%; Pred. No. 2,7e-286;
Matches 689; Conservative 194; Mismatches 305; Indels 34; Gaps 13;
QY 1 MENTQKTVVPTGPGVAVACRVEDDLREISFLAARSTDSLALLPLMNLVETKFTS 60
DB 1 MESAPKTVSLPVSGLGVYAIQNTFMEATLTMAKSIDSLAVLPVINGLVEQTFT 60
QY 61 SLAVVSGARTTGLAGAGITLKLTTSHFVPSVFPVPHGGKVLPSAAAPNLTACNAARERF 120
DB 61 NVAVVAGSKITGLGAGITLKLTPSHFTPNAFYVCGSVFGASKAPNLTACELARRF 120
QY 121 QFSRCQPPVDGAVETTGABICTRLGLEPENTILYLVLTALFKXAVPMCNVFLHYGGLDI 180
DB 121 GFSPFSPVVDNAVETSGEBICASLNLSPETTLVLTVTETFKEMVCMNTFLHYGTST 180
QY 181 VHINHGVDIRLPLFPVQLFMPDVNRLVPDPNTHRSIGEGFVYPTPYNTGLCHLIHDC 240
DB 181 VTTHGQEAVKIPIIPVQLYMPDVNRLAAEPFNKRSIGDEFFVYSKPFNSDJCRLLHY 240
QY 241 VIAPMAVALRVNVTAVARGAAHLAFDENHEGAVLPDITVTYFQSS--SSGTTTARGAR 298
DB 241 VLGPAAVALRVNLDGVARGAAHLALDENHEGAVLPDITVTFLPDSAQGTSGKSGRTOR 300
QY 299 RNDVNSTKSPSGGFERRLASIMAAADTALHAENVTFNTGTYEETPTDIKEWPMFIMGEGT 358
DB 301 QGD-GSGLKGSSSGSERRLASIMAAADTALSVDSIMGAGVYDTLPSSVEDLPI-LSVGDD 358
QY 359 LPRNLALGSYTVARVAVIGAMVFSNSALYLTEVEDSGMTAKOGPGPSNRFVQFAGP 418
DB 359 RERLEALGAYASRLSGLVGAMVFSNSVLTWTEVDDGGPADGKAS-NPSYHLPYLLAAP 417
QY 419 HLAANPQTDREDGHVLS-----SQSTGSSNTEFSDYALICGFGAPILARLLFYLERCD 472
DB 418 YVAGNPQTDKDRVLOHTAQPAPINGSNQFESLDYALACGFCPOLARILFYLERCD 477
QY 473 AGATFG-CHGDALAKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQRPFGQATROP 531
DB 478 AGTTCGNETDALRYLANTLESEVPCGLCTPATRPACAHHTLHRLRQRPFGTTPVAPI 537
QY 532 GVFGTMSQVSDCDPLGNYPYLITRPGDQTEAAKATMODTYRATLERLFDLEGERLL 591
DB 538 GIFGTMNSTVSDCDVLGNVASYGALKRPND-NSAPKSIQMDTYRATMERLVNDLEQAKLI 596

```
592 DRGA-----PCSEGLSSVIVDHPTRRILDLRLARIEQTOTQPMKVLVETRDYKIREGL 646
597 DEALAHAGTCSAS--TGVRKQDQASFINLLSTIKITEGAEEQFVRTLVVRDFKIREGL 654
647 SEATHSMALTPDYGAFCPITNPLVFKTHLAVQDLALSOCHCVFYGOQVEGGRNFRNOF 706
655 ADANTHMSISLDPYSSPCPTSFSLRRTIFAVLQDLVLSOCHCLFYGOQVEGGRNFRNOF 714
707 QVLRFRFVDFNGGFIETSTRTITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVI 766
715 QVLRFRFVDFNGGFIETSTRTITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVI 774
767 RDIRVKRWVSGCNTLSERARLVGLASQYQREKRVDMHLGALGFLIKQFHGLFP 826
775 RDLRIKRWVSGCNTLSERARLVGLASQYQREKRVDMHLGALGFLIKQFHGLFP 834
827 RGVPPNSKSPNPQFWTLQNRQMPADKLTHEETIAAVKRFTEEYAAINFINLPPTCI 886
835 RGHPPGIDTPNPQFWTLQNRQMPARLLSKEDIEITAIKFSHSYSAINFINLTPNI 894
887 GELAQFYMANLILKYCDHSOYLINTLSITIGARRPPSPSSVLHWIRKDVTSADITQA 946
895 GELAQFYANLILKYCDHSOYFINGLTAVVGSRRPRDPAALAWINRTINGASDVEPAA 954
947 KALLEKTENLPELWTTAFTSTHLVRAAMNQPMVVLGISIKYHGAAGNNRVFOAGNWSG 1006
955 QEVLOQLGSGNPAWGTGTASINWVYWDQRPVWVIGLSIKYSGAGNNRVFOAGNWSG 1014
1007 LGGKQNVCLFTFTRRIFIIACPRGGFCIPVTGSSGNRTTSLDQVRGIIVSGGAMVQ 1066
1015 LGGKQNVCLFMAFDRTRRFLVACPRVGTCEAGGFGMGARENTLSEQIRTSIVSDGPMVQ 1074
1067 LAIVATVVRVAGRAQHWAFDDWLSLTDDELARDLELHDQIICLTETPWTVEGLEAV 1126
1075 TAVFSVLTALGARTQHLAVDDWIGLVDDELFLASLDNALNAVVDQF-GEWSVEAAQDMI 1133
1127 KILDEKT-----TAGDGETPTNLAFNEDSCFESHDITTSNVLNISGNSITGTVPLKRRPP 1181
1134 RTMDAQTNWGVVSTGSG-----AFDFGACVDGANQSSSTTFNMGPA--SSSAPAGQKRFH 1185
1182 EDELEFDSGIPKIGNTMMEM 1203
1186 PDDILFDNGAPPEKXSLTFDM 1207

RESULT 3
89549 D Q89549 PRELIMINARY; PRT; 1203 AA.
T T 01-NOV-1996 (TRENBLrel. 01, Created)
T T 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
T T 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
E UL29.
N N
S Bovine herpesvirus 1.
C Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C Alphaherpesvirinae; Varicellovirus.
X NCBI TaxID=10320;
X [1]_
P SEQUENCE FROM N.A.
C STRAIN=COOPER;
A A Schwyzer M., Vlcek C., Lowery D.E., Bello L.J., Meyer G., Misra V.,
A Thiry E., Paces V.;
L Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
P [2]
P SEQUENCE FROM N.A.
C STRAIN=COOPER;
A A Schwyzer M., Vlcek C., Paces V., Pastoret P., Thiry E., Schwyzer M.;
L Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
P [3]
P SEQUENCE FROM N.A.
C STRAIN=JURA;
A A Schwyzer M., Vlcek C., Lowery D.E., Bello L.J., Meyer G., Misra V.;
```

```
Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RC STRAIN=JURA;
RA Schwyzer M.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z98205; CAB01596.1; -
DR EMBL; X94677; CAA64336.1; -
DR EMBL; AJ004801; CAA06104.1; -
DR InterPro; IPR000635; Viral DNA bind.
DR Pfam; PF00747; viral DNA bp. 1
SQ SEQUENCE 1203 AA; 127409 MW; 8299D6456A9654F CRC64;

Query Match 53.6%; Score 3371.5; DB 12; Length 1203;
Best Local Similarity 53.6%; Pred. No. 8.1e-269;
Matches 652; Conservative 197; Mismatches 340; Indels 27; Gaps 16;

QY 1 MENTQKTVVTPGPIGVYVACRVEDLDEEIEFLAARSTDSLDLALLPLMRNLTVKETS 60
DB 1 MDAAAKTVALAGPAGVYVCDAARLDLKALLAARSDSLAVPLVIRGLVETAPAP 60
QY 61 SLAVSGARTGLAGAGITLKLTTSHFVPSVVFHGGKHLVPSAAPNLTRACNARRP 120
DB 61 NYAVVAGTITGLGAGITLKLTPSHYHPNVFVHGGKHLVPSAAPNLTRACNARRP 120
QY 121 GFSRCQGPVDCGAVETGABICTRLGLSPENTILVLTALFKAIVFNCVFLHVGGLD 180
DB 121 GFSAFAGAPVDCGAVETGABICTRAVGSPETALVLTATTEAPKEIVYWCNTFLHYGGA 180
QY 181 VHINHGDIIRIPLFPVQVPMEDVNLVDPDPNTHRSIGEGFVYPTPEYNTGLCHLI 240
DB 181 VQVGAGEAVRVLVPLVQVPMEDVNLVDPDPNTHRSIGEGFVYPTPEYNTGLCHLI 240
QY 241 VIAPNAVALRVNTAVAGAAHLAPDENHAGNLPDITVYTFSSSSGTTTARGARN 300
DB 241 VLGPAVALRVNTLDAVARGAAHLAFDESHAGVLPDPVCFVTF-DQAGAGGGGSGARG 299
QY 301 DVNSTSKPSPSGGFERRLASIMAAOTLHAELVIFNTGIVEETPTDIKEWPMFIMEGL 360
DB 300 APGAKAAPGVVERLLASVMAADTAVSIEAANSTSVFDEDDVACVDDWPMLOGAADA 359
QY 361 RNALGSYTVARVAGVIGAMVFPNSALYLTVEBESGNTAKDGGPSPFNRFYFAGPHL 420
DB 360 KLDALGAVYVRLAGLVGAMVFPSSNSVLTHTVDDGGAADAKDGA-AAAGFHFYQIAFY 418
QY 421 AANPQTRDGHVLSSTG-----SSNTEPSVDYALI CGGAPLALRLFYLERCDAG 474
DB 419 AGNPRCDKGKELPOTGAGPAVINGAQEFALDHLALACGCTQOLLARMLFYLERCDAG 478
QY 475 AFTGGHG-DALKYVTGTDSIETPCSLCEKHTRPVCAHTTVHRLRQRMFRFCOATQPT 533
DB 479 APAGRNMDALAVVASTLEGDVFCGLCSRDDRHACATTLHLRLHRLPRFGAPTRSLGV 538
QY 534 FGTMSQVSDCDPLGNYPYLILRKGQDTEAAKATQDTYRATLERLFDLEERLLDR 593
DB 539 FGTMSAYSDDCVLGNYSYSAIRPG-ADENARSIMQATYRAAVENYLAELDOORLLAH 597
QY 594 GAPCSSEGLSSVIVDHPTRRILDLRLARIEQTOTQPMKVLVETRDYKIREGLSEATH 653
DB 598 DAQSAQAQ-LERAITDHASFGALAAI QNTVQAETAFVIRGLVEDDFQREALYEAHTL 656
QY 654 ALTFDYPYGAFCPTNPLVFKTHLAVQDLALSOCHCVFYGOQVEGGRNFRNOFVLRRR 713
DB 657 SLALDYPYSTAVCPATAFLFRSRSLAVVQDLALSOCHGIFCGQPVDFGRFRAQFOFVLRR 716
QY 714 FVDLFGGFIETSTRTITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRVN 773
DB 717 FMDLLGGFLITRTITVTLSEGPVSAPNPTLGQDAPAGRTFDGDLARVSVEVIRVN 776
QY 774 RVVFGNCTNLSEAPARLVGLASAYQREK-RVMDLHGALGFLIKQFHGLLFRPGMPPN 832
DB 777 RVWFSAGNANMSEAAARVLGLAGAYQKPSGGVNLSPGLGFLVQKPHRLKFLFNKPPG 836
```

QY 833 SKSPNPFQWTLLOQNOPADKLTHEBITTAAVQRTEBYAAINFINLPPTCIGELAQF 892
 DB 837 SPTNPFQWTLLOQNOPADKLTHEBITTAAVQRTEBYAAINFINLPPTCIGELAQF 896
 QY 893 YMANLILKYCDHSOYLINTLTSIITGARRPRDPSPVSLHWIRKDVTSAADETQAKALLEK 952
 DB 897 YMANLILKCDHCKFYINGLTAIVAGARRPRDPAVNVHWVRRRIADAPDAAAABEVLAE 956
 QY 953 TENLPELWTTAFTTHLVRAAMNORPMVLGISISKYHGAAGNRRVFAQNWGLNGGKN 1012
 DB 957 AEARPEYWAGTFAASNLVRSYMATPRAVVLGLSISKYNGSAGNRRVFAQNWGLNGGKN 1016
 QY 1013 VCPLEFTRTRRFFIACPRGFCIPVTGSSGNRETTLSDOVRGII--VSGAMVQLAIYA 1071
 DB 1017 VCPLEFTRTRRFFIACPRGFCIPVTGSSGNRETTLSDOVRGII--VSGAMVQLAIYA 1076
 QY 1072 TVRAVGAQAQHMADFWLSDTDDFLARDLELHDQIIQTLETP--WTVEGALEAVKIL 1129
 DB 1077 AVLQALGPRVHEMDLDDWAALVEDEFFAQSMVELTE---RAAARPGWSPGGAADMLREL 1133
 QY 1130 DEKTTAGDGETPTNLAFNEDSCPSHDTTSNVLN--ISGNSISGSTVPGIKRPPDELEF 1187
 DB 1134 ELEBAABAE--PAGAFDFGACASAPADATYAFGGPVAG---AGPAAPGAKR--PDLELEF 1187
 QY 1188 DLSGIPKHKGNITMEM 1203
 DB 1188 EM-GVPDKRPALTIDM 1202

RESULT 4
 O92611 PRELIMINARY; PRT; 1177 AA.
 ID AC O92611; 1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE DNA-binding protein.
 GN DBP.
 OS Pseudorabies virus.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 OX NCBI_TaxID=10345;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=TNL;
 RX MEDLINE=98455382; PubMed=9784061;
 RA Wu S.-L., Hsiang C.-Y., Ho T.-Y., Chang T.-J.;
 RT "Identification, expression, and characterization of the pseudorabies
 virus DNA-binding protein gene and gene product.";
 RL Virus Res. 56:1-9(1998).
 DR EMBL; U80909; AAC63429.1; -
 DR InterPro; IPR000635; Viral_DNA_bind.
 DR Pfam; PF00747; viral_DNA_bp; 1.
 KW DNA-binding.
 SQ SEQUENCE 1177 AA; 125408 MW; BA87AF9CFC961707 CRC64;

Query Match 52.9%; Score 3327.5; DB-12; Length 1177;
 Best Local Similarity 53.5%; Fred. No. 3.4e-265;
 Matches 655; Conservative 182; Mismatches 318; Indels 69; Gaps 17;
 QY 1 MENTQKTVVPTGPGVYACRVEDLLEBISFLAARSTDSIALPLMNLTVKTFTS 60
 DB 1 MEAAKTVTVRAAPLGYYVTPTEALRRDLALLVARSADDAAVAPLVGLVVEAGFAG 60
 QY 61 SLAVSGARTTGLAGAGITLKTSTHYPVPVPHGKHVLPSSAAPNLTRACNAARERF 120
 DB 61 HVAVVAGARTTGL--GGGLTLKLPNHPNVPFFHNGDCVPPSSAAPALSRACEAARAF 119
 QY 121 GFSCQGFVVDGAVETTGATCIRGLGEPENTILYLWTFALPKEAVPMCNVFLHYGLDI 180
 DB 120 GFSAYR--TFVDNAEETTGAVSCVLAADAAHAAYLVADGPKAEVYLCNAPLHYGGAGT 178
 QY 181 VHINHGDIVRIPLFPVQWLFMPDVNRLVPDPFNTHRSIGSGFVYPTFYNTGLCHLIDC 240

RESULT 5

DB 179 VSIINGEAWRVLPVFLFDPVNVRLVADFNNAKRSISEEYVPRFFENGPCRLHGY 238
 QY 241 VIAPMAVALRVNTAVARGAAHLAFDENEGAVLPDITVTVYFOSSSSGTTTARGARRN 300
 DB 239 VLGPAAVATRVNLDVARGAAHLAFDENESAVLPADVFTTFLFQRRRGDDAA----- 292
 QY 301 DYNSTSKPSPGSGFERELASIMAADTALHAETVINTGTYEETPTDIKEWMPF----- 352
 DB 293 -----AGGLEREMASVMSADAALSLEALVAAGVYDEEPPALDDWVPLSEAGAKDG 342
 QY 353 ---IGMEGTLPRLNALGSYVARVAGVIGAVTSPNSALVLTVEDESGMTEAKDGGPGPS 408
 DB 343 GAAAASAGAPVQAAALGAVVSRRAAGLVGALVSSNSVLYLTVDDDGAADARKEGAGPS 402
 QY 409 FNRFYQFAGPHLAANPQTRDRGHVL-----SSQSTGSSNTSEFSDVYLALICGFGAPLIAL 464
 DB 403 FNRFYQVAAAPYLAGNPQTKDGRVLLHTASOPATAPGNHDFAMDHLMVAGFCFQILLARV 462
 QY 465 LFYLERCDCAGAPTG--GHGDALKVYGTGTFDSEIPLCSICEKTRPVCAHTTVHRLRQMPRE 533
 DB 463 LFYLERCDCAGTFVGRGNDVAVRVVAGSLDAEPVCSICDRASRACAHHTTLHLRHR--SRA 521
 QY 524 QCATRQPTIGVGTWNSOYSDCDPLGNYPALILIRKPGDQTEAAKATMQDTYRATLERLEI 593
 DB 522 SARRRSPMGVGTWNSAYSDDCVLGNVASYGALKRPND--SEPPKAIMQDTYRAAVDRLLA 580
 QY 584 DLEQERLLDRGAPCSSSEGLSSVIVDHPFRRIIDTLRARIEQTTTFQMKVLVETRYKIR 643
 DB 581 DVAGAR-----IGETVDEHAGFPHALRALRDTVEQAAARFVRLTVETRFKIR 628
 QY 644 EGLSEATHSMALTEDPYSGAFCPITNPLVKRTHLAVVQDLALSCQCHVFCVQGVQVEGRNR 703
 DB 629 DALVDANHTWLSLDPISGALCPATSLARPTLLAVLODLALSCQCHGVHGHQPVESGRNR 688
 QY 704 NQOPVLRPRVDFLNGGFTSTRITVTLSEGVSAHPNPFYLGODAPAGRTFDGDLARVSU 763
 DB 689 NQOPVLRPRVDFLNGGFTVAKTIVTVTLADG--IVAPDPTKGSAPPARDHGDGLARVSF 747
 QY 764 EVIRDIVQRVVFSGNCTNLSAARARLVGLASAYORQKRVDMHLHGALGFLLKQFHGL 823
 DB 748 EVLRRLVKSRVMESTGSGLSDAARARVAGLAGAVQRPDTAVDVNLGFLGFLKKOHAT 807
 QY 824 LFPRGMPNKSPPNPQWTLLOQNOPADKLTHEBITTAAVQRTEBYAAINFINLP 883
 DB 808 LFPRGKPPGGSPNPQWTLLOQNOPADKLTHEBITTAAVQRTEBYAAINFINLP 867
 QY 884 TCIGELAQFYMANLILKYCDHSOYLINTLTSIITGARRPRDPSPVSLHWIRKDVTSAADE 943
 DB 868 GTVAELAQFYLANLILKYCDHCKFYINGLTAIVAGARRPRDPAVNVHWVRRRIADAPD 927
 QY 944 TQAKALLEKTENLPFELMTTFTSTHIVRAAMNORPMVLGISISKYHGAAGNRRVFAQNW 1003
 DB 928 RAAREVLDAPRD--DTWVATYTSHELLRSVWASRPVLVGLGVSKYHGWAGNRRVFAQNW 985
 QY 1004 WSLNGKGNVCPLPFTDRTRRFFIACPRGFCIPVTGSSGNRETTLSDOVRGII--VSGAM 1063
 DB 986 WSLNGKGNVCPLPFTDRTRRFFIACPRGFCIPVTGSSGNRETTLSDOVRGII--VSGAM 1045
 QY 1064 MVQLAIYATVAVRARGAQAQHMADFWLSDTDDFLARDLELHDQIIQTLETPWTVEGAL 1123
 DB 1046 LVQTVAVVAUHALGARTQHLBEPDDWRAIVDDEFLAALAEINGRVADR--DGRWSVEAAA 1104
 QY 1124 EAVKILDEKTTAGDGETPTNLAFNEDSCPSHDTTSNVLNISGNSIS--GSTVP--GLKR 1179
 DB 1105 ELVRDLRGQTGADGGGSET--AFDFGACGAGGD-----AGAGLAPASLAPAEELGGKR 1154
 QY 1180 PPEDELEFSLSGIPKHKGNITMEM 1203
 DB 1155 PPEDELEFDMGAPPEK--RLTFDM 1176


```

69101
D Q69101 PRELIMINARY; PRT; 1197 AA.
C Q69101;
T 01-NOV-1996 (TrEMBLrel. 01, Created)
T 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
T 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
E DNA binding protein ICp6.
S Herpes simplex virus (type 2).
S Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C Alphaherpesvirinae; Simplexvirus.
X NCBI_TaxID=10310;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=KN;
X MEDLINE=93228441; PubMed=8385914;
X Toh Y., Tanaka S., Imai Y., Mori R.;
T "Nucleotide sequence of the major DNA binding protein of herpes
T simplex virus type 2 and comparison with the type 1 counterpart.";
L Arch. Virol. 129:183-196(1993).
R EMBL; D10658; BAA01507.1;
R InterPro; IPR000635; Viral DNA bind.
R Pfam; PF00747; viral DNA bd. 1.
Q SEQUENCE 1197 AA; 128470 MW; AA3ADA75B8865BFE CRC64;

Query Match 47.9%; Score 3012; DB 12; Length 1197;
Best Local Similarity 49.0%; Pred. No. 4e-239;
Matches 595; Conservative 212; Mismatches 368; Indels 40; Gaps 16;

Y 4 TQKTVTVTGLGVY--ACRVEDLDEIEISFLAARSTDSALLPLMRNLTVKTTSS 61
b 7 TTTTVKVPVGMVYVGRACPAEGL--LSLLSARSAGDAVAVAPLIVGLTVESGFAN 64
Y 62 LAVVSGARTTGLAGATILKLTTSHPFVSFVFGGKHVLPSSAAPNLTRACNAARERFG 121
b 65 VAAVVGSTTGLGTVASVLSKMPHSYVSFVFGGRLHAPSTQAPNLTRCERARRHF 124
Y 122 FRCGGPVDGAVETTGAEITRIGLEPENTILYLVTALFKEAVFNCVPLVYGGLDIV 181
b 125 FSDVAPRCDLKHTTGDALCERLGLDPRALLYIVITEGPREAVCISNTFLHGGMDKV 184
Y 182 HINQGVIRIPLEFVQLEPMPDNLVPPNTHRSIGEGFVYTPPYNTGLCHLHDCV 241
b 185 TIGDAEVHRIPIPLQMFDPDSRVADPFNCNRSIGENYLPFPNPLARLLFEAV 244
Y 242 IAPMAVALRVNVTAVARGAAHLAFDENHEGALPDPITVYFQSSSGTTTARGARND 301
b 245 VGPAAVALRARNVAVARAHLAFDENHEGALPALDITPTAFASOG--KPGRGAR--- 299
Y 302 VNSTSKPSPSGFERRLASIMAAHTALHAEIFNVTGIYEETPTDIKEWPMFIMEGTLP 361
b 300 --DAGNKGFPAGGFQRLASVMAAGDAALASIVSMVAFDEPPDPITWPLLEGQETPA 357
Y 362 LNALGSYTVAVGVIGAMVFPNSALYLTEVEDSGMTAKDGGPSPFNRFYQFAGPHLA 421
b 358 AGAVGAYLARAGLVAMVFTNSALHTEVDADGADPKDHSK-PSFYRFFLVPGTHVA 416
Y 422 ANQTDGRGHVL-----SSQSTGSSNTEFSVDYALICGFGAPLIARLLFYLERCDAG 474
b 417 ANPQDREGHVVPVGEPRTPAPLVGGTQ-EPAGSHLAMLGFSAPLAKMLFYLERCDGG 475
Y 475 AFTGCHG-DALKYVVTGTFDSIPCSLCEKHTFPCVCAHTVHRLRQRPQATQPIGV 533
b 476 VIVGQEMDVFRYVADSGQTDVPCNLCTFTRHCAHTLMRLRARHPKFAARGAIGV 535
Y 534 FGTMNSQSDCDPLGNVAFYILIRKPGQDTEAAKATWQDTRYATLERLFDLEQERLLDR 593
b 536 FGTMNSAYSDDCVLGNVYAFSAKX-ADGSENTRTIMQETVYRATERVMALEALQYVDQ 594
Y 594 GAPCSSEGLSSVVDVHTPFRILDTLRARIEQTTPQMKVLVETRDYKIREGSEATHSM 653
b 595 AVPTALGLETIIGNREALHTVWNKQVLDREVEQLMRNLIEGRNFKFRDGLAEANHAM 654
Y 654 ALTDFPYSGAFCPITNPLVLRKTHLAVQDLALSQCHCVFQQVQGVQEGNFRNQFVLRR 713

```

```

Db 655 SLSDDYTCGCPCLQLLARRSNLAVYQDLASUQCHGVFAGQVEGRNFRNQFQVLR 714
QY 714 FVDLFGGFISTRSITVTLSEG-PVSAPNPTIGQDAPAGRTFDGLARVSVEVIRDIRV 772
Db 715 VMDLPNNGFLSAKTLTVALSEGAACAPSLTAGQAPAESFEGDVARVTLGFPKELAVK 774
QY 773 NRVVSGNCTNLSEARARLVGLASAYQORQKRVMDLHGALGFLKQFHGLLFFRGMPPN 832
Db 775 SYVLFPAGASANASEAKARVASLQAYQKPKRVDILLGLGLFLKQFHAVIFPNKPPG 834
QY 833 SKSPNPQFWTLLQRONQPADKLTHEEITTTAAVKRFTTEVAAINFILNPLPTCIGEL 892
Db 835 SNQPNQFWTALQNLQPARLSREDIETAFIKRFSLDYGAINFINLAPNNSVSELY 894
QY 893 YMANILKYCHDSOYLINTLSITGARRPDPSPSVLHWIRKDVTSAAIDTQAKALLEK 952
Db 895 YMANQILRYCHDSHTVFINTLTAVIAGSRPPSPVQAAAAPQ--GGAGLEAGARALMDS 951
QY 953 TENLPELWTTAFTSLVRAAMNQPMVVLGISIKYHGAAGNRRVFOAGNWSGLNGGKN 1012
Db 952 LDAHFGATSMFASCNLLRPVMAAPMVVLGLSISIKYGMAGNDRVFOAGNWSLLGKN 1011
QY 1013 VCPLETFDTRRFFIIACPRGGFICPTGPPSSGNRETTLSDDQVGIIVSGAMVQIAIYAT 1072
Db 1012 ACPELLIFDTRKFLACPRAGFVCAASSLGGGAHESLCEQLRGIIAEGGAIVASSVFA 1071
QY 1073 VVRANGARQAHVAFDDWLSLTDDEFLARDLELHDQIIQTLETPTWVEGAL-----EAVKI 1128
Db 1072 TVKSLGPRTQQLQIEDWLALLLEDEYLSEMMFETTRALERGHGEMSTDAALEVAHEAL 1131
QY 1129 LDEKTTAGGTGPTNLAFNPOSCEPSHDTTNSVLNLSGNTSGSTVPGCLKRPPEDDELFD 1188
Db 1132 VSQLGAGE-----VNFQDGFQDDEDDHAASFGGLAAA-AGAAGVAKRAFAFGDDFFG 1183
QY 1189 LSGIPIKGNITMEM 1203
Db 1184 -EGPPEKK-DLTLD 1196

RESULT 6
Q9PZ52 PRELIMINARY; PRT; 1190 AA.
AC Q9PZ52;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Homolog of HSV-1 glycoprotein.
CS Turkey herpesvirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Marek's disease-like viruses.
OX NCBI_TaxID=10390;
RN [1]_TaxID=10390;
RP SEQUENCE FROM N.A.
RC STRAIN=HPR24;
RA Kato K., Jiang H., Izumiya Y., Cai J., Teushima Y., Miyazawa T.,
RA Kai C., Mikami T.;
RT "Identification and Transcriptional Analysis of the Marek's Disease
RT Virus Serotype 2 Genes Homologous to the Glycoprotein B (UL27), the
RT ICPI8.5 (UL28) and the Major DNA-binding Protein (UL29) Genes of
RT Herpes Simplex Virus Type 1."
RL J. Vet. Med. Sci. 0:0-0(1999).
DR EMBL; AB024711; BAA83753.1;
DR InterPro; IPR000635; Viral DNA bind.
DR Pfam; PF00747; viral DNA bd. 1.
SQ SEQUENCE 1190 AA; 130339 MW; D9291E2EAA7F59D CRC64;

Query Match 44.4%; Score 2796.5; DB 12; Length 1190;
Best Local Similarity 45.3%; Pred. No. 2.6e-221;
Matches 546; Conservative 235; Mismatches 396; Indels 29; Gaps 13;

QY 6 KTVTVPTPLGVVYACRVEDLDEIEISFLAARSTDSALLPLMRNLTVKTTSSLA 65

```


Db 654 LAVPYASGICPVAFUSRTTIAVQDMALSCSIVMOQCVQVARNFRFTQFOAVLKRVR 713
 QY 715 VDLFNGGFISTRSITVTLSGCPVAPNPTLQDAPAGRTFDGLARVSVEVIRDIRVKR 774
 Db 714 LELQVAGFITSKTTITVLEQQICVDPDSQVSDSVSNWEGDLVKVTVVEIFELVKVK 773
 QY 775 VPSGNCNTN-LSEARARLVGLASAYQQRKRVDMHLGALGFLFKQPHGLLPFRGMPPNS 833
 Db 774 VLFQGGIAGAASEATKSRLAGVVEAYQRPKTHVLNGLPGLGFAVKSYHTLLFPDVVRPVG 833
 QY 834 KSNPQWFTLLQRONMPADKLTHEEITTTAAVKRPTTEEVAAINFNLPTTCIGELAQFY 893
 Db 834 ATPNALWFWILLNQLPAGILSKEEDKSLFKTKTSYADWNYINISPTCPGDLAQFY 893
 QY 894 MANILKYCDHSQVLYINTLSITIGARRPRDPSVHLWIRKOVTSADITQAKALLEKT 953
 Db 894 LANILKYCSHKHFFINTISALVAVSRPRDPAIVLPFIERPITKGDVAPAAQLIASM 953
 QY 954 ENLPFWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNRRVFOAGNWSGLANGKV 1013
 Db 954 SDKDIWCAFTSINLGSIMTKPFVVICISISKYHGMAGSTKVQSGNWNMGGRNV 1013
 QY 1014 CPITFDTRFRFIACRGGFICVPTGPSSGNRETTISDOVRGIIIVSGGAMVQLAIYATV 1073
 Db 1014 CSLMSFDRTHRYVMTCPRVGFVAEQPIFSSGIXKSTTLIDRVMLVSEESAAPHAAVYMLA 1073
 QY 1074 VRAVGAQAHMADWLSLTDDEFLARDLEELHDOIIQILETPTWVEGALEAVKILDEKT 1133
 Db 1074 LKMGVDRVQMELEDMWEIINDEYISLDELKQ-VVEAGGNADAAMTLAKEMVNA 1132
 QY 1134 TAGDGETPTNLAFNFCSPESHDTTNSVNLINSGNISGTVPGLKRPDDDELFDLSGIP 1193
 Db 1133 MSIPTDGPT---PFDACDENLGHADGQTISETN-----LKP--NMNVFLEPIP 1179
 QY 1194 IK 1195
 Db 1180 EK 1181

RESULT 9

29DPQ8 QDQP8 PRELIMINARY; PRT; 1190 AA.
 ID QDQP8
 AC QDQP8
 JT 01-MAR-2001 (T-EMBLrel. 16, Created)
 JT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
 JT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
 DE UL29 single stranded DNA binding protein.
 3N HVT037.
 CS Melesgrid herpesvirus 1 (herpesvirus of turkeys).
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 CC Alphaherpesvirinae; Marek's disease-like viruses.
 CX NCBI_TaxID=37108;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FC126;
 RA MEDLINE=20578232; PubMed=1134310;
 RX Afonso C.L., Tulman E.R., Lu Z., Zaak L., Rock D.L., Kutish G.F.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RT "The genome of turkey herpesvirus";
 RL J. Virol. 75:971-978(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FC126;
 RA Afonso C.L., Tulman E.R., Lu Z., Zaak L., Rock D.L., Kutish G.K.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF291866; AAC45767.1;
 DR InterPro; IPR000635; Viral_DNA_bind.
 DR Pfam; PF00747; Viral_DNA_bp_1;
 DR NCBI_TaxID=37108;
 SQ SEQUENCE 1190 AA; 130054 MW; B450CAPSFFFD928 CRC64;

Query Match 44.2%; Score 2781.5; DB 12; Length 1190;
 Best Local Similarity 44.5%; Pred. No. 4.4e-220;
 Matches 538; Conservative 245; Mismatches 391; Indels 35; Gaps 14;

QY 1 MENTKQTVTPTGKLVYACRVEDDLERISFLAARSTSDSLALLPLMNLNVEKFTFS 60
 Db 1 MDGVKSKVLGGPIGVVYNSKDAVPDELSTFAAKCIDSESSILPLIAGLVESEFVR 60
 QY 61 SLAVVSGARTTLAGAGITILKLTSSHYPVSVFPHGKRVLPSSAAPNLTRACNAARERF 120
 Db 61 NVAAVAGTKTKLVLNGGITILKLVTHYHNVFVYGGDCIKPCSKAPNLTKACEYARERF 120
 QY 121 GFSRCQPPVDGAVETTTGABICTRLGLEPENTILYLVTALFKEAVFMCNVFLHYGGLDI 180
 Db 121 GYCYSAP---GSVETSGKQICEQLGLDAQKMLYLVTFLKEAYLNCNLFYGGSDS 177
 QY 181 VHINGVDIRPLPPVOLFWPDMVNRVLPDPNTHRSIGEGFVYPTFFYNTGCHIHDC 240
 Db 178 VTINDAEVRIRPLPLHLVLPDFNRVSEPFSEPRPALGDGAYMPEAFFNDALCRLLGY 237
 QY 241 VIAPMAVALRVNVTAVARGAHLADENHEGAVLPDITYTYE---QSSSGGTTARGAR 298
 Db 238 VLGTDAVGLVRNIDAVARGAHLCPDENHEGILLPADTFTTADTGGKGQPKAGK 297
 QY 299 RNDVNSTKPSGPFERRLASIMAADTALHAEVIFNTGIYEETFTDIKEWPMFIGMEGT 358
 Db 298 RDGAE----TQGGYERRTASLMASDATTIAENVIASSVVEEPIPDVKKWPIFCNPVG 352
 QY 359 LPRNALGSTARVAGVIGAMVSPNSALYLTVESDSGTEAKDGG-PQPSNRFYQFAG 417
 Db 353 TDRVEALSAYNGRVAGLVGAMVFNSSNVIYMTVEGSSSDGKESGVPAFPYRFFQIAA 412
 QY 418 PHLAANPQTDQRDGHVLSQS---TGSSNTEFSDVLYALICGPGAPLALRLLFYLERCDA 473
 Db 413 PHLAANPLVDRDCKPISGDALPKLAPSTSEYSLDVLILAGFCPLLARFLFYLERCDG 472
 QY 474 GAFTHGHG-DALKVYOTTFDSEIIPCSICEKHTRPVCAHTTVHRLRORMPRFGQATQPIG 532
 Db 473 GANACHDLDLVKLVSADIDAMPCELCDKTISVYCAHTTVKLEVLKPKFGVQMGAMG 532
 QY 533 VFGTWNQSDCDPLGNVAPYLILRKPGDQTEAAKATMODTYEATLERIFIDLEQERLLD 592
 Db 533 LFGMTWNYCDVNLGSAQVSTLKR--SEGEASRSVMQDTYLTVERVWKALEKEGLLA 590
 QY 593 KGAPCSSEGLSSVIVDHTFRILDLRLARIEQTTPMKVLVETEDYKIREGLSEATHS 652
 Db 591 CDDPMNMAPADASIKDQPSFIHAVSTWNRNIEGEASQLARNLTREYINIREGLCDANHT 650
 QY 653 MALTFDPYSGAFCTNFIKRVTHLAVQDLALSQCHVYCOOQVGRNFRNQFQVLR 712
 Db 651 LSLTIEPSSGFCVLSFLARRTTIAVQMALSCSMLMHGQGVQVARNFRFTQFQSVLR 710
 QY 713 RFVDLFNGGFISTRSITVTLSGCPVAPNPTLQDAPAGRTFDGLARVSVEVIRDIRVK 772
 Db 711 RVLELQVAGFITSKDTITATLEDOHVAVPDPSPQYDPTVINMEGDLTVTIELRELKVK 770
 QY 773 NRUVFSGNCT-NLSEARARLVGLASAYQQRKRVDMHLGALGFLKQPHGLLPFRGMPP 831
 Db 771 NRUVFSGNCT-NLSEARARLVGLASAYQQRKRVDMHLGALGFLKQPHGLLPFRGMPP 830
 QY 832 NSKSPNQPWFILLQRONMPADKLTHEEITTTAAVKRPTTEEVAAINFNLPTTCIGELAQ 891
 Db 831 NGAVPNAHWFTLLQRONLPDMLSKEDQKALFIKFTNVYADWNYINISPTGFDLIAQ 890
 QY 892 FYWANLILKYCDHSQVLYINTLSITIGARRPRDPSVHLWIRKOVTSADITQAKALLE 951
 Db 891 FFPLANTILKYCSHKHFFINTISALVAASKKPRDPAIVLPWIDKTTITQGRDVSIAAOLIG 950
 QY 952 KTELNPBLWTTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNRRVFOAGNWSGLANGKV 1011
 Db 951 TIADRKDAWCAFTSINLGSIMTKPFVVICISISKYHGMAGSTKVQSGNWNMGGRNV 1010
 QY 1012 NVCEPLFTDTRFRFIACRGGFICVPTGPSSGNRETTISDOVRGIIIVSGGAMVQLAIYA 1071
 Db 1011 NVCSLMSFDRTHRYVMTCPRVGFVAEQPIFSSGIXKSTTLIDRVMLVSEESAAPHAAVY 1070
 QY 1072 TVRAVGAQAHMADWLSLTDDEFLARDLEELHDOIIQILETPTWVEGALEAVKILDE 1131


```

Db 1071 LALKWGERVRQWELDEWITDDTSSIAADLNEQAQS-EGGWSIDVAMGLAK---- 1125
Qy 1132 KTAGGETPTN-LAFNFDSCPSHDTTNSVLNIGSNGISGTVPLKRPEDDELFDLS 1190
Db 1126 ETAAASAFIPSGPTTNYBEWDDGVKERSD-----GHSISEG---GTRKP--NLSVFDLD 1175
Qy 1191 GPIKHGNI 1199
Db 1176 PVPEKRAAV 1184

RESULT 10
>9E1G3
ID Q9E1G3 PRELIMINARY; PRT; 1190 AA.
AC Q9E1G3
DT 01-MAR-2001 (T-REMBLrel. 16, Created)
DT 01-MAR-2001 (T-REMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (T-REMBLrel. 19, Last annotation update)
DE UL29 major DNA binding protein.
RN UL29.
RS Meleagrid herpesvirus 1 (herpesvirus of turkeys).
SC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Marek's disease-like viruses.
OX NCBI_TaxID=37108;
RN [1]
IN
PC SEQUENCE FROM N.A.
STRAIN=FC126;
TX MEDLINE=21195611; PubMed=11297687;
TX Kingham B.F., Zelnik V., Kopacek J., Majerciak V., Ney E.,
TX Schmidt C.J.;
TX "The genome of herpesvirus of turkeys: comparative analysis with
TX Marek's disease viruses.";
TX J. Gen. Virol. 82:1123-1135 (2001).
TX EMBL; AF282130; BAG30069.1;
TX InterPro; IPR000635; Viral_DNA_bind.
TX Pfam; PF00747; Viral_DNA_bp; 1.
Q  SEQUENCE 1190 AA; 129952 MW; 4A419239FB47DD82 CRC64;

Query Match 43.9%; Score 2764.5; DB 12; Length 1190;
Best Local Similarity 44.3%; Pred No. 1.1e-218;
Matches 536; Conservative 245; Mismatches 393; Indels 35; Gaps 14;

Y 1 MENTOKTVPTGPGYVYACRVEDLDEEISFLAARSTDSLALLPLMRNLTVKEFTS 60
b 1 MDGVGSKVLGGPIGYVYSKDAVPDELSTFAAKCIDSEESTLPLIAGLTVESEFVR 60
Y 61 SLAVSGARTTGLAGAGITLKLTTSHFVPSVFVHGKHVLPSSAAPNLTRACNAARERF 120
b 61 NVAAVAGTKTGLVNGGTTKLVPVTHYHPNVFVYGGDCIRPCSKAPNLTKACEYARERF 120
Y 121 GFSRCQGPVDAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHVGLDI 180
b 121 GVCPSAP---GSVETSKQICEQLGLDAQKTMVLTWTELFKEAIVLCNSFLYGGSDS 177
Y 181 VHINQGVIRPLFPVQLFMDVNRVLPDPNTHRSIGEGFVYPTPPYNTGLCHLIHDC 240
b 178 VTINDAEVRRIPLYPLHLVLPDFNRVSNPEFSERPRALGDGAVNPEAFNDALCRLLGY 237
Y 241 VIAPMAVALRVNTAVARGAHLAFDENHGVLPDITVYTP--QSSSGTTTARGAR 298
b 238 VLGTDAVGLRVNTDAVARGAHLCFDENHGLIPLPADTTTATPTADTGGKQPKAGK 297
Y 299 RNDVNSTSKPSPSGGFERRLASIMAADTALHAEIFNTGIYEETPTDIKEWPMFIMEGT 358
b 298 RDGAE-----TQGGYERTASLMASDATLAIENVIASVVEEIPDVKKWPIFCNVGY 352
Y 359 LPRNLALGSYARVAGVIGAMVFPNSALYTEVEDSGTEAKDGG-PGSPFNRYFOFAG 417
b 353 TDRVEALSAYMGRVAGLVGAMVFPNSVIYNTVEGEASSDGSKGSGVPAPSFYRFFQIAA 412
Y 418 PHLAANPOTDRDGHVLSQS-----TGSSNTEFFSDYLLALICFGAPLILALLFLYLRCD 473

```

```

Db 413 PHLAANPLVDRDGGKPSGDALPKLAFSTSTSEYSLDVLILACGFCPOLLARFLFYLRCDG 472
Qy 474 GAFTGGHG-DALKVVTGTFDSEIPCSLCEKHTPEVCAHTTVHRLRORMPRFGQATROPIG 532
Db 473 GANACHDLDTVKLVSSAIDADMPCELCDSKYCAHTTVKLEVLKPKFGVOMRGAMG 532
Qy 533 VFGMNSQYSDCPLGNAPYLILRKPGDQTEAAKATMODTYATLERLFDLEQERLLD 592
Db 533 LFGANTNNYCDVNALGSAQFSLKE--SEGEASRSVMODTYELTVRYVMKALEKEGLLA 590
Qy 593 RGAPCSSEGLSSVLDHPTERRILDLRARIEQTTFQFMKVLVETDYKIREGLSEATHS 652
Db 591 CDDPMWAPADASIKGASFIHAVSTMRIIEGASQLMRNLTEIRESNIREGLCDANHT 650
Qy 653 MALTFDYPGAFCPITNLFVKRTHLAVQDLSLSQCHVYGCQVEGRNFRNQFQVLRAR 712
Db 651 LSLTIEPYSGFCVLSFLARRTTIAVIODMALSQCSMLMHGQVNEARNFRTOFQSVLR 710
Qy 713 RFVDLNGGPISTRTSVTLSEGPVSAPNPTLQDAPAGRTTFDGLARVSVVEIRDIRVK 772
Db 711 RVLELQAGPITSKDTATLEDOHVAVDPDSQYDPTVINMEGDLTRVTIIRIELKV 770
Qy 773 NRVVFGNCT-NLSEARARLVGLASAYQQRKQVMDLHGALQFLIKQFHLFFPRGMP 831
Db 771 NRVIFGGVIGTASDAKSLANWIEAYQPTKMWNLNGPLGPAIKVYHSALFPNVKVP 830
Qy 832 NSKSPNQFWTLQONQPADKLTHERITTAAVKRFTEYAAINFINLPTCIGELAQ 891
Db 831 NGAVPNAHFWTLQONOLPDTMLSKEDQKALFKKFTNVYADNMVYNISPTGFDLAQ 890
Qy 892 FYANLILKYCHSOYLINTLSITIGARRPDPSSVLHWIRKDVTSADIDTQAKALLE 951
Db 891 PFLANTILKYCHKHFFINTISALVAAKPKRDPALVLPWIDKTTQGRDVSIAAQQLIG 950
Qy 952 KTELPELMTTFTSTHLVRAANQRMVYLGISIKYHGAAGNNRVFQAGNWSGLNGK 1011
Db 951 TIADRKDAWCATFSATNLVGSVMSTKPLIVIGVSIKYHGMAGSTKVFGSGNWMNGR 1010
Qy 1012 NYCPLTFTDTRFPIACRGGFCVTPGSPGNRETTILSDOVRGIIVSGGAMVQLAIYA 1071
Db 1011 NVCSLMSFTRTHRYMACPRVGFVAHTGFSGLKEATLMRARAILEEGGAPHAEVYM 1070
Qy 1072 TVVRVAGAAQHAQFDDMLSLTDDFLARDELHDIQTLETPTWTEGALAEVKILDE 1131
Db 1071 LALKWGERVRQWELDEWITDDTSSIAADLNEQAQS-EGGWSIDVAMGLAK---- 1125
Qy 1132 KTAGGETPTN-LAFNFDSCPSHDTTNSVLNIGSNGISGTVPLKRPEDDELFDLS 1190
Db 1126 ETAAASAFIPSGPTTNYBEWDDGVKERSD-----GHSISEG---GTRKP--NLSVFDLD 1175
Qy 1191 GPIKHGNI 1199
Db 1176 PVPEKRAAV 1184

RESULT 11
>9E1B6
ID Q9IBU6 PRELIMINARY; PRT; 1191 AA.
AC Q9IBU6
DT 01-OCT-2000 (T-REMBLrel. 15, Created)
DT 01-OCT-2000 (T-REMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (T-REMBLrel. 19, Last annotation update)
DE UL29.
RN UL29.
RS Turkey herpesvirus.
SC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Marek's disease-like viruses.
OX NCBI_TaxID=10390;
RN [1]
IN
PC SEQUENCE FROM N.A.
STRAIN=CA;
TX MEDLINE=92237304; PubMed=1315048;
TX Jones D., Lee L., Liu J.L., Kung H.J., Tillotson J.K.;

```


"Marek disease virus encodes a basic-leucine zipper gene resembling the fos/jun oncogenes that is highly expressed in lymphoblastoid tumors.";
 Proc. Natl. Acad. Sci. U.S.A. 89:4042-4046 (1992).
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-GA;
 RX MEDLINE=20283955; PubMed=10823954;
 RA Lee L.F., Wu P., Sui D., Ren D., Kamil J., Kung H.J., Witter R.L.;
 RT "The complete unique long sequence and the overall genomic
 RT organization of the GA strain of Marek's disease virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:6091-6096 (2000).
 DR EMBL; AF147805; AAF66764.1;
 DR InterPro; IPR000635; Viral_DNA_bind.
 DR Pfam; PF00747; Viral_DNA_BP.1.
 SQ SEQUENCE 1191 AA; 130946 MW; AC04A0158B0797A3 CRC64;

Query Match 43.1%; Score 2710; DB 12; Length 1191;
 Best Local Similarity 44.1%; Pred. No. 3,6e-214;
 Matches 535; Conservative 236; Mismatches 400; Indels 42; Gaps 16;
 QY 1 MENTQKTVTPGPGYVYACRVEDLDEEISF-----LAARSTDSDALLPLMRNLTV 54
 DB 1 MDGVGSKVLGGPIA-IYMER-----PNVAFPMWNCDLAAKSNDCDDAVLPIVAGLTV 53
 QY 55 EKTFTSSLAUVSGARTTGLAGAGITKLTTSHFPVSFVFGGHVLPSSAANPLTRACN 114
 DB 54 ESDFWNVAAGVTKTGLSGGGTTLKLVPHYHPCVFFVGGDCIKPCTKAPNLTKACD 113
 QY 115 AARERFGFSRCQGPVDCGAVETTGAEICTRLGLEPENTILYLVVTLFKEAVFMCNVFLH 174
 DB 114 LARERFGVS-AVSPAPATPTTQGCCLEALMDAQNVMYLVVTLFKEVILNSFLH 172
 QY 175 YGGLDIVHNGDVIRPLFPVQVLPMDPNRVLDPDPNTHRSIGEGVYPTPFYNTGLC 234
 DB 173 FGGSDVNTINNADVRPIYPLHVLDPFNRTNTEPFSEKPRALGEGAIKPKAFYNDSLC 232
 QY 235 HLIHDCVIAPMAVALRVNVTAVARGAAHLAFDSNHEGAVLPDDITVYFQSSSGTTTA 294
 DB 233 RLLHGYVLTAVGLRVNIDAIARGAAHLCFDNHEGTLIPADTTFTATTPAE--TT 289
 QY 295 RGARRNDVNSTKSPSGGFFRRRLASIAADTALHAEIFNTGIVETPTTDIKWPMFIF 354
 DB 290 KGQSKMGKREGSDVS-GGGYERRTASLWASDATLAIENWISASVYEDPIPDVKNKWPYCN 348
 QY 355 MEGTLPRINALGSAVAVGIVGAWFSPNSALVLTSEVDSGMTXAD-CGPGSPFNRFY 413
 DB 349 PVGVADRIEALUSATWAVAGLVGAWFSSNVIYMTSEVGEAGSTEGTETSTTAPSYRFP 408
 QY 414 QFAGPHLAANPQTRDRGHLVS---SQSTGSSNTEFSDVYLALICGAPLILARLLFYLE 469
 DB 409 QIAAAPHLSANPLVDRDGPVSGENLSKSTASQSEYSLDYLILACGFCPOLLARFLYLE 468
 QY 470 RCDAGATGGHG-DALKVYGTGTFDSEIPCSLCEKHTFPCVCAHTVHRLORMPRFQATR 528
 DB 469 RCDGGAKASHDLDTVRFVSAIDADMPCLCDKTSRIYCAHTTIKRLVRLPKFYQWR 528
 QY 529 QPIGVFGTMSQYSDCDPLGNAPYLILRKPQDTEAAKATMDQTYRATLERLFDLEOE 588
 DB 529 GAMGLFGAMTNNYCDNALVSQAQFTLKR--SEGEASRSVMQDTYRLTVRMWMALEKE 586
 QY 589 RLLDRGAPCSSEGLSVSVVDPHTPRRLDITLARIETTTQFMKVLVETRDYKIREGL-S 647
 DB 587 GLLTCDPTNMAADANIRDGNFSPIRAISTMKNIESEASQLMNETIREVNTITGVNG 646
 QY 648 EATHSMALTDPYSGACPTINFVAKETHLAVQDIALSOCHCVFYGOQVGENPFRNQF 707
 DB 647 DANHTLSLAVEPYASGICPVLAFLSRITIIAVQDMALSOCSIMQOQQVBARNFPTQF 706
 QY 708 PVLRREVDLPNGGFISTRSITVLTSEGPVSAPNPTLQDAPAGRTFDGDLARVSVEVIR 767
 DB 707 AVLKRRVLELQNAQFITSKITVLTLEDOQICVDPSPSKSQYDSVISNMGEDLVKVTIEFR 766

QY 768 DIRVQRVVFSGNCTN-ISEAAARLVGLASAYQOEKRVDMHLGALGFLKOPHGLLFP 826
 DB 767 ELKVXKVLFGGAGIAGASEATKSLAGWYAYQPTTMMVNLGPGLOFAVKRYHTLLFP 826
 QY 827 RGMPPNKSQPNQFWTLQORNPADKLTHEITITIAAVKRFTEEYAAINFILPPTCI 886
 DB 827 DYKMPNGATPNALWFILLRNLQNPAGILSKEEEDKSLFIKFTKSYADMNVINISPTCF 886
 QY 887 GELAQPYMANLILKYCDHSOYLINTLTSITGAREPRDPSVVLHWIRKDVTSADIETQA 946
 DB 887 GLDAQPYLANTLILKYCSHKHFFINTISALVAASRAPRDPALVLPWIERPIYKQGVAPAA 946
 QY 947 KALLEKTEMPELWTAFTTSTHLVRAAQNQPMVVLGISISKYHGAAGNRRVFOAGNWSG 1006
 DB 947 QQLIASMSDHKIMCATPSSINLAVGSIWTKFVFVIGISIKYHGMAGSTKVFQSGNWN 1006
 QY 1007 LGGKNVCPFTFTDTRFIACRGGPICVTPGSSGNRETTLSDOVRGIIVSGGAMVQ 1066
 DB 1007 IMGGRNVCSLMGFDTRHYVMTCPRVGFVAQPIFSSGIKETTLIDRVVVLSESAAPH 1066
 QY 1067 LAIYATVTVAVGARAQNAFDDWLSLTDDFLARDLELHQIITQTLTPTWVEGALEAV 1126
 DB 1067 AAVYMLALKWGDVRVQWELEDWMEITNDEYISLIDELNKQ-VEEAEGGWNAADAAMTLA 1125
 QY 1127 KILDEKTTAGDGETTNLAFNDFSCPSHDSTSNVLNIGSNISSGTVPGUKRPPEDDEL 1186
 DB 1126 KEMVNMAMSIPTDGT---FDFDADENLEGHADGQITSETN-----LKRDP--NNNV 1172
 QY 1187 FDLGIPIKHGNI 1199
 DB 1173 FDLPIPEKRAPV 1185

RESULT 12
 Q9QH63 PRELIMINARY; PRF: 999 AA.
 AC Q9QH63; 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE DNA binding protein.
 OS Gallid herpesvirus 1.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Infectious laryngotracheitis-like viruses.
 OX NCBI_TaxId=10386;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SA-2;
 RA Johnson M.A.;
 RT "Sequence of the infectious laryngotracheitis virus (SA-2 strain)
 RT unique long region UL28 to UL43".
 RL Submitted (JUL-1999), to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF168792; AAD56201.1;
 DR InterPro; IPR000635; Viral_DNA_bind.
 DR Pfam; PF00747; Viral_DNA_BP.1.
 SQ SEQUENCE 999 AA; 108977 MW; 089FAED1C8BD674B CRC64;

Query Match 23.0%; Score 1445; DB 12; Length 999;
 Best Local Similarity 30.5%; Pred. No. 8.7e-110;
 Matches 351; Conservative 177; Mismatches 376; Indels 246; Gaps 18;
 QY 13 GFLGVYACRVEDLDEEISFLAARSTDSDALLPLMRNLITVEKTTSSLAUVSGARTT 71
 DB 22 GFIGVYVAPREEDGKEWKLCAKSDQPSCAIPLIGLTVESDFKPNVAAVIGTKSS 81
 QY 72 GLAGNGITLKTTHSFVPSVVFHGGKHVLPSSAANPLTRACNARERFGFSRC-QGPPV 130
 DB 82 GVWGNCNTALSPCHPFTTYVIFHGSECIPPTSTFNTLKICEDARERFGFSSLPNGPV 141
 QY 131 DGAVETTGAEICTRLGLEPENTILYLVVTLFKEAVFMCNVFLHGYGLDIVHNGDVIR 190
 DB 142 PNAETTGEEICKSLANDPDKTMLYLVVAEPFCEAVVYCNVTHYGSVETLYINSKHVR 201

191 IPLPVPQLEWPDVN-RLVDPDPNTHRSIGEGFVYPTPFYNTGLCHLHDCVIAVAL 249
 202 IPIVPMYMDIALRICRNPDSNENIGECIYPTPLFNKALNRLHATLGTQGS 261
 250 RVNVTAVARGAAHLAFDENHGAVALPPIITVYFSSSSGTTTARGARNVNSTKPS 309
 262 RTNDEAVARGAAMLAFCGSGFCGLTSDKTTEL-----AQNIQQTGP 307
 310 PSGGFERRLASIAAATAALHAEIFNFTGTYESTPTD-----IKEMPMFIGNEGLPLRN 363
 308 QNVEVERRAACSLAELATLISVCAPY---PFDNGGLPYEWEPLFCCKTQWERVQ 364
 364 ALGSYTARVAGVIGAMVGFNPGNLSALYLTVEYDSGMTAKDGGSG-----PSFNRYFOAGPH 419
 365 ALDRFSAELAGVGAIFNPGNLSALYLTVEYDSGMTAKDGGSG-----PGERASSGLTRFYLTCLH 419
 420 LAANPOTDRGH-----VLSSOSTGSSNTEFSDVIALICGFGAPLLARLLFYLERCDAG 474
 420 LIGCQVDVSGKRVNVEGTCPSPLTQNTFEYDPEHLAYACGSPSELLSRIIFYLEKSKQ 479
 475 AFTGGHGDALKYVTGTFTFSEIPCS-----LCEKHTRPVCAHTTVHRLRORMPRFGQAT 527
 480 LSTREN---MNLISINSKCACTGAVPECRKCNDESNNHCIRYTMQRLTRFPFPRMTR 536
 528 ROPIGVFTGTMNSQYSDCDPLGNVAYLILRKPDQTEAAKATMDQTYRATLERLIDLEQ 587
 537 RSPIAVIGAVDGYTDCDILGNFAYSQLKAGD-GEPAKAVMNDYRGIGWRVFOYLV 595
 588 ERLDRGAPCSSEGLSSVIVDHTFTRILLTLARIEOTTTCPMKVIVETDCKIREGLS 647
 596 EGLINKDGTEDTHNLSL-----AEKTVFEKIONFVGSECKFISALUSGVRAYHYKSHLS 651
 648 EATHSMALTFFPYGAFPCPTNPLNKRTHLAVQDLALSQCHVYFQQVQGRNFRNQFQ 707
 652 STVHTFG-SPEPYSSAYCPVLSLLCAQTRSLFDLILSLQIHGTFDTRQPSKMFSSAM 710
 708 PVLRERFVDFLNGGISTRSITVILSEGVSAHPNPTLQODAPAGETDGDGLARVSEVIR 767
 711 PALRAFNGMLDKGSLGKYEFPFSISVPCVNAPDTPRENTQCAIQAQVEYLSRQGVILKX 770
 768 DIRKRVVVFSGNCTNLSEAAARLVLASAYQKQKVDMLHGLALGFLKQPHGLLPPR 827
 771 EFKVKNRVFVNGSSFSNS----- 788
 828 GMPNSKSPNQFWTLQORNQMPADKLTHEEITIAAVKFTTEYAAINFILPPTCIG 887
 789 ----- 788
 888 ELAQYMANLILKYCDHSQYLINTLTSITGARRPRDPSSVHLWIRKDVTSAADETQAK 947
 789 ----- 788
 948 ALLEKTENLPBLWTTAFTSTHLVRAANNQRMVVLGISISKYHGAAGNRRVFOAGNWSGL 1007
 789 -----CRVLSKFPVLMGIMISKYVQONSSTTVFQSGNWSL 826
 1008 ---NGKNVCPLEFTRRRIIACRGGFTCPVTPGSSGNRETTLSQDVRGIIVSGAM 1064
 827 IGNSGQSNVSNALNDPVRKFAKACKRGVITLSSGQVASCQFQENTIASQVRSIIDAGGS- 885
 1065 VQAIATVAVRAVGAQAQMAFDWLSLTDDEFLARDLEELHDOIIQOTLETPWTVEGALS 1124
 886 -PTSITVAVLVGEGMKDITAEITWMAITDDKYLINILVELREDISGS-QNGSVATAQS 943
 1125 AVKILDEKTT 1134
 944 MLSELEGKDT 953

RESULT 13
 40913
 040913
 040913;

PRELIMINARY;

PRT; 1132 AA.

DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE ORF 06.
 OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Rhadinovirus.
 OX NCBI_TaxID=37296;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97296220; PubMed=9151804;
 RA Neipel F., Albrecht J.C., Fleckenstein B.;
 RA "Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus
 human herpesvirus 8: determinants of its pathogenicity";
 RL J. Virol. 71:4187-4192(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Neipel F., Albrecht J.-C., Ensser A., Huang Y.-Q., Li J.J.,
 RA Friedman-Kien A.E., Fleckenstein B.;
 RA "The genome of human herpesvirus 8 cloned from Kaposi's sarcoma";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U93872; AAB26603.1;
 DR InterPro; IPR002114; Hpr Serp site.
 DR InterPro; IPR001993; Mitoch carrier.
 DR InterPro; IPR000635; Viral DNA_bind.
 DR Pfam; PF00747; Viral DNA bp. 1
 DR PROSITE; PS00215; MITOCH_CARRIER; 1.
 DR PROSITE; PS00589; PTS_HPR_SER; 1.
 SQ SEQUENCE 1132 AA; 125394 MW; CF8BE4F451919CBF CRC64;
 Query Match 10.9%; Score 683.5; DB 12; Length 1132;
 Best Local Similarity 23.5%; Pred. No. 9e-47;
 Matches 301; Conservative 196; Mismatches 500; Indels 285; Gaps 53;
 QY 11 PTGPIGVYACVREDLDEEISFLAARSTDSLLPLMRLNLTVEKTTSSLAVVSGART 70
 DB 18 PTGCPGVLAYLTHFPICEASLLGNGYPEAKVFSPLHLGLTVESDPLNVKAVH--- 73
 QY 71 TGLAGAGITLKTTHSHFVSVFVHGXHVLPSAANPLTRACNAARERFGRSCQGPV 130
 DB 74 KIDATASVKTLSH---REALVFNTHLFPFGKLEKLCRESLRFSTFVEQOH 131
 QY 131 DGAVETTGABICTRIGLEPENTILYLVVLTALFKEAVFNCVFLHYGGLDI-----VH 182
 DB 132 KGTL--WSPACQPLPCANE-IFMAVIVTEGKERL-----YGGKLVFVPSQTPVH 180
 QY 183 INHGDVIRIPLFPVOLFMVDVNLVDPFNTHRSIGEGFVYPTPFYNTGLCHLHDCVI 242
 DB 181 IGEHQAFKILPYDEDLFGPSRAQEL-----CRFYNPDISRYLHDSIF 222
 QY 243 APMAVALRVNVTAVARGAAHLAFDENHGAVALPPIITVYF-QSSSSGTTTARGARN 301
 DB 223 TGIQAALRVKDVSTVIQ-ASERQF--VHDQYKIKLVQAKDPQOCASRGTDG----- 271
 QY 302 VNSTKSPSGGFERRLASIAAATAALHAEIFNFTGTYESTPTDIKE-----WPMFIG 354
 DB 272 -----STLMVID-SLVAELCMGYGLSFIEGPQDSCEVLNVDTPWIFEN 313
 QY 355 MBGTLPRLNALGSYTARVAGVIGAMVGFNPGNLSALYLTVEYDSGMTAKDGGSPFNFYQ 414
 DB 314 CETPPARLALAEVHAEQALHIGALFAANSVLVLTTRV--AKUPQKNQORDANNYSFYL 371
 QY 415 FAGPHLAANPQTRDGG-----HVLSSOSTGSSNTEFSDVIALICGFGAPLLARLLFYLE 469
 DB 372 QHGLGYLEATVKGASAFKGVFVSALDGGSS---YTLQHLAYASSFSPHLLARMCYLQ 428
 QY 470 RCDAGFTGGHG-DALKYVTGTFTFSEIPCSCEKHTRPVCAHTTVHRLRORMPR-FQAT 527
 DB 429 FLEHHKNTNSQSYNVVDYV-GTAAPSQMCDLCCQCCPACVNCINTLFRMKDRFPFVLSNVK 487
 QY 528 ROPIGVFTGTMNSQYSDCDPLGNVAYLILRKPDQTEAAKATMDQTYRATLERLIDLEQ 587
 DB 488 RDPYVITGTACT-YNDLEILGNLFPATFREREEGNVEADP---KYTTWQLCQNI-----T 538

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

DM protein - protein search, using sw model

Run on: January 30, 2004, 13:07:07 ; Search time 17 Seconds
(without alignments)
3327.831 Million cell updates/sec

Title: US-09-769-699-2

Perfect score: 6294

Sequence: 1 MENTQKTVTPGPGYVYA.....DELFDLSGPIKHGNTWEM 1203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6294	100.0	1204	1 DNEI_VZVD	P09246 varicella-z
2	3620.5	57.5	1209	1 DNEI_HSVB	P28932 equine herp
3	3054	48.5	1196	1 DNEI_HSVK	P17470 herpes simp
4	3049	48.4	1196	1 DNEI_HSVL	P04296 herpes simp
5	3047	48.4	1196	1 DNEI_HSVI	P17469 herpes simp
6	3015.5	47.9	1196	1 DNEI_HSVH	P89452 herpes simp
7	3013	47.9	1197	1 DNEI_HSV2	P36384 herpes simp
8	2933	46.6	1186	1 DNEI_HSV2	P12639 bovine herp
9	1094.5	17.4	375	1 DNEI_HSV21	Q03444 equine herp
10	669	10.6	1128	1 DNEI_EBV	P03227 Epstein-Bar
11	604	9.6	1128	1 DNEI_HSV6A	P24910 herpesvirus
12	492.5	7.8	1132	1 DNEI_HSV6U	P52338 human herpe
13	490.5	7.8	1132	1 DNEI_HSV6Z	P52538 human herpe
14	483	7.7	1131	1 DNEI_HSV7J	P52339 human herpe
15	479.5	7.6	1160	1 DNEI_SCMVC	P13215 simian cyto
16	477	7.6	1235	1 DNEI_HCMVC	P17147 human cytom
17	473	7.5	1193	1 DNEI_HSV72	Q09177 herpesvirus
18	412	6.5	1191	1 DNEI_MCMV2	Q36717 murine cyto
19	121	1.9	1066	1 Z295_HUMAN	Q01133 homo sapien
20	115	1.8	823	1 CRK3_CHICK	P18461 gallus gall
21	114	1.8	822	1 IF2C_EUGER	Q09417 caenorhabdi
22	114	1.8	1941	1 YRM8_CAEL	P16654 rattus norv
23	113.5	1.8	1886	1 GP21_EAT	P12785 rattus norv
24	113	1.8	2505	1 FAS_RAT	Q61789 mus musculu
25	110	1.7	3333	1 LNA3_MOUSE	Q61315 mus musculu
26	109	1.7	2845	1 APC_MOUSE	Q60106 xanthomonas
27	108	1.7	827	1 XANP_XANS2	Q02470 lactobacill
28	108	1.7	1902	1 P2P_LACPA	P88937 pseudomonas
29	107	1.7	1279	1 BGSC_PSEBL	Q06J90 homo sapien
30	105.5	1.7	3298	1 PC16_HUMAN	Q03132 saccharopol
31	105.5	1.7	3567	1 ERY2_SACER	P15558 pseudomonas
32	106	1.7	773	1 PAC2_PSSS3	P03186 epstein-bar
33	106	1.7	3149	1 TSGU_EBV	

RESULT 1

ID	DNEI_VZVD	STANDARD;	PRT;	1204 AA.
AC	P09246;			
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Major DNA-binding protein.			
GN	29			
OS	Varicella-zoster virus (strain Dumas) (VZV).			
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;			
OC	Alphaherpesvirinae; Varicelloviruses.			
OX	NCBI_TaxID=10338;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86306857; PubMed=3018124;			
RA	Davison A.G., Scott J.E.			
RF	"The complete DNA sequence of varicella-zoster virus";			
RL	J. Gen. Virol. 67:1759-1816(1986).			
CC	-I- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA REPLICATION.			
CC	-I- SUBCELLULAR LOCATION: Nuclear (Probable).			
CC	-I- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN FAMILY.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

CC	EMBL; X04370; CAA27912.1; .
DR	PIR; C27214; DNB29
DR	InterPro; IPR000635; Viral_DNA_bind.
DR	Pfam; PF00747; Viral_DNA_bp; 1.
KW	DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT	ZN FING 497 510 C2HC-TYPE
SQ	SEQUENCE 1204 AA; 133139 MW; D2FEFE65DC0CC674 CRC64;

Query Match 100.0%; Score 6294; DB 1; Length 1204;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MENTQKTVTPGPGYVYACRVEDLDLEISFLAARSTDSDLAALLPLMRNLTVKEFTFS	60
DB	1	MENTQKTVTPGPGYVYACRVEDLDLEISFLAARSTDSDLAALLPLMRNLTVKEFTFS	60
QY	61	SLAVVSGARTTGLAGAGITLKTTSHPYSPVFPHGKHYLPSSAANLFRACVAARERF	120
DB	61	SLAVVSGARTTGLAGAGITLKTTSHPYSPVFPHGKHYLPSSAANLFRACVAARERF	120
QY	121	GFSCRQGPVVDGAVETTGAEICTRLGLEPENTILYLVVTALFKFAVFCNVFLHYGLDI	180
DB	121	GFSCRQGPVVDGAVETTGAEICTRLGLEPENTILYLVVTALFKFAVFCNVFLHYGLDI	180

09/769,699 212/04
Search Notes


```
181 VHINHGDIIRPLPPVLFMPDVRNLPDPNTHRSITGEGVYVPTPYNTGLCHLIHDC 240
181 VHINHGDIIRPLPPVLFMPDVRNLPDPNTHRSITGEGVYVPTPYNTGLCHLIHDC 240
241 VIAPMAVALRVNTAVARGAAHAFDENHEGAVLPDITTYTFQSSSGTITARGARN 300
241 VIAPMAVALRVNTAVARGAAHAFDENHEGAVLPDITTYTFQSSSGTITARGARN 300
301 DVNSTKSPSGGFERELASMAADTALHAEVIFNTGIYEETDIDKEWPMFIMGSTLP 360
301 DVNSTKSPSGGFERELASMAADTALHAEVIFNTGIYEETDIDKEWPMFIMGSTLP 360
361 RLNALGSYTRAVAGVIGAMVFPNSALYLTEVEDSGMTEAKDGGPSPNRFYQFAGPHL 420
361 RLNALGSYTRAVAGVIGAMVFPNSALYLTEVEDSGMTEAKDGGPSPNRFYQFAGPHL 420
421 AANPOTDRDGHVLSQSTGSNTSEFSDYIALICGFGAPILAKLLFVLRCDAGATGGH 480
421 AANPOTDRDGHVLSQSTGSNTSEFSDYIALICGFGAPILAKLLFVLRCDAGATGGH 480
481 GDALKYVTGTFDSEIPCSLCEKTRPVCAHTTVHRLRORMPRFGQATROPIGVFGTMNSQ 540
481 GDALKYVTGTFDSEIPCSLCEKTRPVCAHTTVHRLRORMPRFGQATROPIGVFGTMNSQ 540
541 YSCDPLGNVAPYLILKPGQTEAAKATQDQYRATLERLFDLQERLLDRGAPCSSE 600
541 YSCDPLGNVAPYLILKPGQTEAAKATQDQYRATLERLFDLQERLLDRGAPCSSE 600
601 GLSIVVDHFTFRILDTLBARIEQTTQPMKVLVETRDYKIREGLSEATHSWALTFDPI 660
601 GLSIVVDHFTFRILDTLBARIEQTTQPMKVLVETRDYKIREGLSEATHSWALTFDPI 660
661 SGAFCPITNPLVKTTHLAVVQDLALSCCHVCFYQQVQVEGNFRNQPVLRFRFVDFLNG 720
661 SGAFCPITNPLVKTTHLAVVQDLALSCCHVCFYQQVQVEGNFRNQPVLRFRFVDFLNG 720
721 GFISTSIITVLTSEGPVSNPNTLGDAPAGRTFDGLARVSVEVIRDRVKNRVVFSGN 780
721 GFISTSIITVLTSEGPVSNPNTLGDAPAGRTFDGLARVSVEVIRDRVKNRVVFSGN 780
781 CTNLSEAAARLVGLASAVORQEKRVDMHLHGALGFLKKQFHGILLPRGMPNPKSPNPQW 840
781 CTNLSEAAARLVGLASAVORQEKRVDMHLHGALGFLKKQFHGILLPRGMPNPKSPNPQW 840
841 FWTLLORNQMPADKLTHEEITITAAVKRTEEVAAINFNLPTTCIGELAQFYMANLILK 900
841 FWTLLORNQMPADKLTHEEITITAAVKRTEEVAAINFNLPTTCIGELAQFYMANLILK 900
901 YCDHSQYLINTLTSITGARRPRDPSVLHWIKOVTSAADETQAKALLEKTEMLPELW 960
901 YCDHSQYLINTLTSITGARRPRDPSVLHWIKOVTSAADETQAKALLEKTEMLPELW 960
961 TTAFTSTHLVRAAMNORPMVLGISISKYHGAAGNNRVFQAGNWSGLNGKNCVCLFTFD 1020
961 TTAFTSTHLVRAAMNORPMVLGISISKYHGAAGNNRVFQAGNWSGLNGKNCVCLFTFD 1020
1021 RTRERFIACPRGGFCVPTGSPSGNRETTLSQVRGIIIVSGGAMVQLAIYATVRAVGAR 1080
1021 RTRERFIACPRGGFCVPTGSPSGNRETTLSQVRGIIIVSGGAMVQLAIYATVRAVGAR 1080
1081 AQHMAFDMLSLTDDDEFLLARDLEELHDQIIQTLETPTWTEGALEAVKILDEKTTAGDGET 1140
1081 AQHMAFDMLSLTDDDEFLLARDLEELHDQIIQTLETPTWTEGALEAVKILDEKTTAGDGET 1140
1141 PTLNLFNFSCEBPSHDTTGNVLNIGSNTSGSTVGLCKRPBDDBELFDLSGIPKHKGNIT 1200
1141 PTLNLFNFSCEBPSHDTTGNVLNIGSNTSGSTVGLCKRPBDDBELFDLSGIPKHKGNIT 1200
1201 MEM 1203
1201 MEM 1203
```

```
RESULT 2
DNBI HSVB
ID DNBI HSVB STANDARD; PRT, 1209 AA.
AC P28932;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major DNA-binding protein.
GN 31.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses, dsDNA viruses, no RNA stage, Herpesviridae,
OC Alphaherpesvirinae, Varicellovirus.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9229566; PubMed=1318606;
RT Telford E.A.R., Watson M.S., McBride K., Davison A.J.,
RT "The DNA sequence of equine herpesvirus-1",
RL Virology 189:304-316 (1992).
CC -1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: M86664; AAB02466.1;
CC PIR: E36798; DNEC4;
CC InterPro: IPR000635; Viral DNA bind.
CC Pfam: PF00747; Viral DNA bp, 1.
CC DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
CC ZN FING 503 516 C2HC-TYPE.
CC SEQUENCE 1209 AA; 129982 MW; 1A728FB0448APE95 CRC64;
Query Match 57.5%; Score 3620.5; DB 1; Length 1209;
Best Local Similarity 57.1%; Pred. No. 4.7e-270;
Matches 698; Conservative 191; Mismatches 299; Indels 35; Gaps 14;
QY 1 MENTKTVTVPTPLGVYACRYEDLDLEEISPLAARSTDSILALPLMRNLTKERTTS 60
DB 1 MESAPKTVSLPVSPLGVYARQASLOTGTSLTAARSVDSLAVLPVIRGLTVGQTFIT 60
QY 61 SLAVSGARTTLAGAGITLKLTTSHFYPSPVFFHGKHLFSSAAPNLTRACNAARERF 120
DB 61 NVAVVAGSKTGLGGTGITLKLTPSHFNENAFVYGGSVIGASSNAPNLTRACEAARRF 120
QY 121 GSRCCQPPVDGAVETTGAEICTRLGLSPENILVYVLTALKEAVFCNVFLHVGGLDI 180
DB 121 GFAPSSPPVENAVTSGEIEICASINLSPETLALVLTVESFKEMYVCTNLFHYGGTST 180
QY 181 VHINHGDIIRPLPPVLFMPDVRNLPDPNTHRSITGEGVYVPTPYNTGLCHLIHDC 240
DB 181 VTIDGQDAMKIPYVQVLPVDPVNLASEPFFNAKHSIGDERFVSRPFNSDLCHLHCY 240
QY 241 VIAPMAVALRVNTAVARGAAHAFDENHEGAVLPDITTYTFQSSSGTITARGARN 298
DB 241 VLGPAAVALRVNLDGARGAAHLALDENHEGSLVLPQDFTFLPDTQONAGKSGRAQR 300
QY 299 RNDVNSTKSPSGGFERELASMAADTALHAEVIFNTGIYEETDIDKEWPMFIMGST 358
DB 301 QGD-GSGSKNSASSGIERELASVMAADTALSVDSTWAGIYDTHELPSVEDWELSSGDDT 359
QY 359 LPRINALGSYTRAVAGVIGAMVFPNSALYLTEVEDSGMTEAKDGGPSPNRFYQFAGPH 418
DB 360 -ESLEALGAYAAARLSGLVGAMVFSANSLVYMTVEDDGGPADGKGS-NPSVHRFYLIAAP 417
```


Db 530 RGAIGVGTGMSYSDCDVLGNAAAFSAKLR-ADGSETARTIMQETVRAATERVMAELET 588
 Qy 588 ELLIDRGAPCSSEGLSSVVDHPTRILDTLRARIEQTTFQMKVLVETRDYKIREGLS 647
 Db 589 LQYVDQAVPTAMGLETTITNREALHTVNNVROVDREVEQLMRNLVEGRNFKFRDGLG 648
 Qy 648 EATHSMALTFDPSGACFPITNFIKTHLAVQDLALSOCHCVFVQGVQVQVGNERNQFO 707
 Db 649 EANEHMSITLDPYACGCPCLQLLGRSRLNLAQDLALSOCHGVAFQGVQVGNERNQFO 708
 Qy 708 PVLRRFVLDLNGGFISTRSTVTLSEG-PVSAFNPTLGDAPAGRTFDGDLARVSVEVI 766
 Db 709 PVLRRFVLDLNGGFISTRSTVTLSEG-PVSAFNPTLGDAPAGRTFDGDLARVSVEVI 768
 Qy 767 RDIVKRVVFSNGCTNLSAARBARLVGLASAVORQKRVDMHLGALFLLKQFGLLFP 826
 Db 769 KELRVKSEVLFAGASASAEAKARVASQAKPKRVDIILLGLPLKQFHAATP 828
 Qy 827 RGMPPSKSPNPQFWTLQORNPADKLTHEITIAAVKRTVEYAAINFILNPTCI 886
 Db 829 NGKPPGNSQNPQFWTLQORNPADKLTHEITIAAVKRTVEYAAINFILNPTCI 888
 Qy 887 GELAQFYMANLILKYCHDSOYLNTLTSITGARRPRDPSVLHWIRKQVTSADTETO 946
 Db 889 SELAMYMANOILRYCHDSYFINTLTAIGSRPPSVQAAAAW---SAQGGAGLEAGA 945
 Qy 947 KALLETENLPELMTTFTSTHVLRAAMNORPMVILGISISKYHGAAGNRRVFOAGNWS 1006
 Db 946 RALMDAVIDHAGWTSFASCNLLRPWMAAPMVVILGISISKYHGAAGNRRVFOAGNWS 1005
 Qy 1007 LMGKQKVCPLFTDRTRFIIACPRGFCVPTGSPSGNRRTLSQVGRGIIVSGAMVQ 1066
 Db 1006 LMGKQKVCPLFTDRTRFIIACPRGFCVPTGSPSGNRRTLSQVGRGIIVSGAMVQ 1065
 Qy 1067 LATVATVRAVAGARAOHMAFDDMLSLTDDLEFLARLEELHDLQIOTLETPWTEGAL 1123
 Db 1066 SSVFVATVKSIGRPTQOLEDMLLEDBLSEEMWELTARALRNGEWSHTDALEVA 1125
 Qy 1124 -EAVKILDEKTAGDGETPTNLAFNFD--SCEPSHDTTGNVLNISGNSISGTVPLKLEP 1180
 Db 1126 HEAEALVSLQGNAGE-----VFNGDFCE-----DDNATPFGGPGAPGAPAGRKEA 1173
 Qy 1181 PEDELFDLSIGIPIKHNITWEM 1203
 Db 1174 FHGDDPFG-EGPPDKKGDLLDM 1195

RESULT 4

DDBI_HSV11 STANDARD; PRT; 1196 AA.
 AC P04236
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Major DNA-binding protein (infected cell protein 8) (ICP 8 protein).
 DN DBP OR UL29 OR ICP8
 OS Herpes simplex virus (type 1 / strain 17).
 SC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 XC Alphaherpesvirinae; Simplexvirus.
 DX NCBI_TaxID=10299;
 RX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88274327; PubMed=2839594;
 RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
 RA McNab D., Perry L.J., Scott J.E., Taylor P.;
 RA "The complete DNA sequence of the long unique region in the genome of
 RA herpes simplex virus type 1";
 RA J. Gen. Virol. 69:1531-1574(1988).
 RP [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86067223; PubMed=2999714;
 RA Quinn J.F., McGeoch D.J.;
 RA "DNA sequence of the region in the genome of herpes simplex virus

RT type 1 containing the genes for DNA polymerase and the major DNA
 RT binding protein.";
 RL Nucleic Acids Res. 13:8143-8163 (1985).
 RN [3]

RP SEQUENCE OF 1062-1196 FROM N.A.
 RX MEDLINE=88306232; PubMed=2457278;
 RA Hamerschmidt W., Conraths F., Mankertz J., Buhk H.-J., Pauli G.,
 RA Ludwig H.;

RT "Common epitopes of glycoprotein B map within the major DNA-binding
 RT proteins of bovine herpesvirus type 2 (BHV-2) and herpes simplex
 RT virus type 1 (HSV-1).";
 RL Virology 165:406-418(1988)

CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
 CC REPLICATION.

CC -!- SUBCELLULAR LOCATION: NUCLEAR. IN THE ABSENCE OF DNA REPLICATION,
 CC FOUND IN THE NUCLEAR FRAMEWORK-ASSOCIATED STRUCTURES

CC (PERREPLICATIVE SITES); AS VIRAL DNA REPLICATION PROCEEDS, IT
 CC MIGRATES TO GLOBULAR INTRANUCLEAR STRUCTURES (REPLICATION

CC COMPARTMENTS).
 CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN

CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; X14112; CAA32322.1; -
 CC EMBL; X03181; CAA26940.1; -
 CC EMBL; M21631; AAA45787.1; -
 CC PIR; A03790; DNEBVL

CC Interpro; IPR000635; Viral DNA_bind.
 CC Pfam; PF00747; viral DNA_bp.1

CC DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
 KW ZN FING 499 512 C2HC-TYPE

FT SEQUENCE 1196 AA; 128349 MW; 453799162E5B99B9 CRC64;
 SQ

Query March 48.4%; Score 3049; DB 1; Length 1196;
 Best Local Similarity 49.8%; Pred. No. 4.4e-226;
 Matches 609; Conservative 199; Mismatches 367; Indels 48; Gaps 17;

Qy 1 MENTQKIVT--VPTGFLGVY--ACRVEDLDERISFLAARSTDSDLALPLMRNLTYE 55
 Db 1 MEYKPKTATIKVPPGFLGVYARACPSGIEL--LALLSARSGSDVAVAPLVGLTYE 58

Qy 56 KTFSSLAIVSGARTTLAGAGITLKTTSHPYPSYVFPVHGKHLVLPSSAAPNLTRACNA 115
 Db 59 SGFEANVAVVGSRTTGLGTAVSLKLTSPSHYSSVYVPHGGRHLDPSTQAPNLRLCER 118

Qy 116 ARERFGPSRCQGPVVDGAVETTGAEICTRLGLEPENTILYLVVLTALFKAEVPCNVLHY 175
 Db 119 ARRHFGSDVTPRGDLKHETTGELCEGLDLPDRALLYLVVTEGKEAVCINTFLH 178

Qy 176 GGLDIVHINHGDIIRIPLEFPVQLFMPDVNRLVDPDPNTHRSIGRGVYPTFYNTGLCH 235
 Db 179 GGSCKVTIGGAEVHRIPVFLQFLFMPDFSKVLAEFPNANRHSIGKFTYPLFFNRLNR 238

Qy 236 LIHDCVIAPIAVARVNTAVARGAAHLAFDENHGAIVLPPDIITYTTFQSSSSGTTAR 295
 Db 239 LLEFAVVGPAVALRCRNVDAVARAAHLAFDENHGAALPADIITFAFASQG--KTTPR 296

Qy 296 GARRNDVNSKPSGSGGFERRLASITMAADTALHAETVNTGIVETPTDIDKEWPMELGM 355
 Db 297 GGR----DGGKGAAGGFEQRLASVMAGDAAALAESIVSMVAFDEPPTDIDISAWPLFQ 352

Qy 356 EGTLPRLNAGSTARVAGVIGAMVSPNSALYTEDESGMTAKGGPGSPFNRYOF 415
 Db 352 DTAARANAVGAYLARAAGLVGAMVFSNSALHLETVDDAGPADPKHKK-PSFYRFFLV 410

Qy 416 AGPLAANPQTDGHDHVL-----SSQSTGSSNTFSVDYLALICGFCAPILALLFYL 468

```

db 411 PGTHVAANFQVDRGHVVPFGFEGRETAPLVGGTQ-EFAGEHLAMLGFSALLAKMLFYL 469
y 469 ERCDAGFTGGHG-DALKYVTGTFDFSEPCSCERKHTPVCATHTVHLRQMRPFQAT 527
b 470 ERCDGAVVGRQEMDFRVVADSNQDVPNCUCTFDTHACVHTTLMELRARHPKFSAA 529
y 528 RQPIGVFGTMMNSQYSDCDPLGNAPYLILRKPGDTEAAKATMQDTYRATLERLFDLEQ 587
b 530 RGAIGVFGTMMNSQYSDCDVLGNVAFSALKR-ADGSETARTIMQETVRAATERVMAELET 588
y 588 ERLLDRGAPCSSEGLSSVVDHPTFRRLDTRARIEQTTFQMKVLYVETDYKIREGLS 647
b 589 LQYVDQAVPTAMGRLETTITNREALHTVNNRVQVDRVEQLMRNLVEGRNFRDGLG 648
y 648 EATHSMALTDFPYSGAFCPITNELVKETHLVAVQDIALSOCHCVFYGOQVEGGRFNQFQ 707
b 649 EANHANSLLTDFYACGCPPLQLLGRNSNLAVQDIALSOCHGVFAGQSVGGRFNQFQ 708
y 708 PVLRRFVLDLNGGFISTRITVTLSEG-PVSAFNPTLGGDAPAGRTFDGLARVSVEVI 766
b 709 PVLRRFVLDLNGGFISTRITVTLSEG-PVSAFNPTLGGDAPAGRTFDGLARVSVEVI 768
y 767 RDIRYKRVWFSNGCNLSSEARARLVGLASAYOROEKRVDMHGLGALLKQPHGLFP 826
b 769 KELRYKRVLVFAGASANASBAKARVASLQAYQKPKRVDILLQPLGLKQPHAAIFP 828
y 827 RGMPPNSKSPQWFTLQONQPADKLTHEITIAAVKRFTTEYAAINFINLPPTCI 886
b 829 NGKPPGNSQNPQWFTALQONQLPARLLSREDIETIAFKKFSLDYGAINFINLPNV 888
y 887 GELACFYMANILKYCDHSQYLNTLSIIIGARRPRDPSVLHWIRKDVTSADIETQA 946
b 889 SELAMYNANOILKYCDHSTFTINTLTAIGARRRPPSVQAAAAW---SAQGGAGLEAG 945
y 947 KALLEKTENPELMTTFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNKNRVFOAGNWSG 1006
b 946 RALMDAVIDAHPCAWTSMFASCNLLRPVMAARPVMVVLGLSISKYHGAAGNKNRVFOAGNWS 1005
y 1007 LGGKNCVCLPTERTRRFTIACRGFIQVTPGPPSGNRETTLSQVRGIIIVSGGMWQ 1066
b 1006 LMGGKNCVCLPTERTRRFTIACRGFIQVTPGPPSGNRETTLSQVRGIIIVSGGMWQ 1065
y 1067 LAIYATVVRVARGAQAQWAFDMLSLTDDFLARDLEELHDQIIQTLETPTWTEGAL--- 1123
b 1066 SSVFVATVKSIGPTQQLQIEDWLALLEDEVLSEMMELTARALERNGENWSTDALEVA 1125
y 1124 -EAVKIIDEKTTAGGTPTNLAFND--SCEPSHDTTSNVLNISGNSIGSTVPGIKRP 1180
b 1126 HEABALVSQNLNAGE-----VENFGDFGCE-----DDNATPFGGPGAPGAPAGKRA 1173
y 1181 PEDDELFDLSGIPTKHGNITWEM 1203
b 1174 FHGDDPFG-EGPPPKKGLDLDLM 1195

```

RESULT 5

DNI_HSVIF

DNI_HSVIF

STANDARD;

PRT; 1196 AA.

P17469;

01-AUG-1990 (Rel. 15, Created)

01-AUG-1990 (Rel. 15, Last sequence update)

16-OCT-2001 (Rel. 40, Last annotation update)

Major DNA-binding protein (Infected cell protein 8) (ICP 8 protein).

DBP OR UL29 OR ICP8.

Herpes simplex virus (type 1 / strain F).

Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

Alphaherpesvirinae; Simplexvirus.

NCBI TaxID=10304;

[1]

SEQUENCE FROM N.A.

MEDLINE=88306231; PubMed=2841793;

Hammerschmidt W., Conraths F., Mankertz J., Pauli G., Ludwig H.,

Bukh H.J.;

"Conservation of a gene cluster including glycoprotein B in bovine

herpesvirus type 2 (BHV-2) and herpes simplex virus type 1 (HSV-1).";

Virology 165:388-405(1988).

-/- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA

REPLICATION.

-/- SUBCELLULAR LOCATION: NUCLEAR. IN THE ABSENCE OF DNA REPLICATION,

(PREREPLICATIVE SITES); AS VIRAL DNA REPLICATION PROCEEDS, IT

MIGRATES TO GLOBULAR INTRANUCLEAR STRUCTURES (REPLICATION

COMPARTMENTS).

-/- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN

FAMILY.

PIR; D29242; DNEBEF.

InterPro; IPR000635; Viral DNA_bind.

Pfam; PF00747; viral DNA_bp_1.

DNA-binding; DNA replication; Zinc-finger; Nuclear protein.

ZNFING 499 512

SEQUENCE 1196 AA; 128373 MW; BC872584DDB1C8E2 CRC64;

Query Match

Best Local Similarity 49.8%; Score 3047; DB 1; Length 1196;

Matches 609; Conservative 199; Mismatches 367; Indels 48; Gaps 17;

QY 1 MENTQKVT---VPTGPLGYVY--ACRYVEDLDLEBISFLAARSTDSDLALLPLMNLTFVE 55

Db 1 METKPKTATTIKVPPGPLGYVYARACPSEGIEL--LALLSARSQSDVAVAPLVGLTVE 58

QY 56 KFTSSLVAVSGARTTGLAGITILKLTSTHPYSVFPVHGKVLVPSAAPNLTRACNA 115

Db 59 SGFEANVAVVSGRTTGLGTAVALSLKTPSHYSSVSVVFGHRLDPSQAPNLTRLCER 118

QY 116 ARERFGRCQCPFPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKEAVFMCNVFLHY 175

Db 119 ARRHGFSQDYTPRCDLKHETTGALCELGLDPRALLYLVTGFEKVAISINTFLHL 178

QY 176 GGLDIVHNGDVTRIPFLPVLQFMPOVNRVLPDPFNTHRSIGEGFYVPPFYNTGLCH 235

Db 179 GGSQKVTIIGAEVHRIPYVPLQFMFPDSRVIAFPFANHRISIGEKPTYPLPFENRPLNR 238

QY 236 LTHDCVIAPMAVALRVNRVAVARGAAHLAFDENHGAFLPDPITYTYFQSSSGTGTAR 295

Db 239 LLEAVVQPAVALRCNRVDVAVAAAHAFDENHGAALPADITFTAFASQG--KTPR 296

QY 296 GARRNDVNSTKSPSPGFFERRLASIMAADTALHAEIFNTGIYEETPTDIKEWPMFIGM 355

Db 297 GER-----DGGKGAAGFEQRLASVMAGDAALAESIVSMAVDEPPTDISAMPLEFGQ 351

QY 356 ECTLPLNALGYSYARVAGVIGAMVFPNSALYLTVEVDSQMTAKGCGPGSPNRYQF 415

Db 352 DTAARANAVGAYLARAGLVGAMVFTSNLSALHTEVDDAGPADPKDHSK-PSFYRFLV 410

QY 416 AGPHLAANPQTRDRGHVL-----SSQSTGSSNTEPFDYALICGFGAPLARLLFYL 468

Db 411 PGTHVAANFQVDRGHVVPFGFEGRETAPLVGGTQ-EFAGEHLAMLGFSALLAKMLFYL 469

QY 469 ERCDAGFTGGHG-DALKYVTGTFDSEIPCSLCKHTRPVCAHTTTHRLRQMRPFQAT 527

Db 470 ERCDGAVVGRQEMDFRVVADSNQDVPNCUCTFDTHACVHTTLMELRARHPKFSAA 529

QY 528 RQPIGVFGTMMNSQYSDCDPLGNAPYLILRKPGDTEAAKATMQDTYRATLERLFDLEQ 587

Db 530 RGAIGVFGTMMNSQYSDCDVLGNVAFSALKR-ADGSETARTIMQETVRAATERVMAELET 588

QY 588 ERLLDRGAPCSSEGLSSVVDHPTFRRLDTRARIEQTTFQMKVLYVETDYKIREGLS 647

Db 589 LQYVDQAVPTAMGRLETTITNREALHTVNNRVQVDRVEQLMRNLVEGRNFRDGLG 648

QY 648 EATHSMALTDFPYSGAFCPITNELVKETHLVAVQDIALSOCHCVFYGOQVEGGRFNQFQ 707

Db 649 EANHANSLLTDFYACGCPPLQLLGRNSNLAVQDIALSOCHGVFAGQSVGGRFNQFQ 708

QY 708 PVLRRFVLDLNGGFISTRITVTLSEG-PVSAFNPTLGGDAPAGRTFDGLARVSVEVI 766

```

Db 709 PVLRVVDMFNGSLAKTLTVALSEGAACASLTAGQTAPABSSFGDVARVTIGFP 768
QY 767 RDIRVKNRVSGNCTNUSEAARLVGLASAYQOEKRVDMHGLGFLKQPHGLLFP 826
Db 769 KELRVKSVLFGAGASANAESAARVASLQAYQKPKRVILLGLGFLKQPHAAIFP 828
QY 827 RGMPPNSKPNQFWTLQONQPADKLTHEBITTAAVKRPTEEVAAINFILPTCI 886
Db 829 NGKPPSGNQPNQFWTLQONQPADKLTHEBITTAAVKRPTEEVAAINFILPTCI 888
QY 887 GELAQFYMANILKYCHDSOYLINTLSIITGARRPRDPSSVLHWIRKDVTSADIETO 946
Db 889 SELAMYYMANQILRYCHDSOYLINTLSIITGARRPRDPSSVLHWIRKDVTSADIETO 948
QY 947 KALLEKTENLPELWTAFTSTHLVRAAMQPMVVLGISIKHGAAGNNRVFOAGNWSG 1006
Db 946 RALVDAVDAHFGAMTSPFASCNLLRPVMAARPMVVLGLSISKYGMAGNDRVFOAGNWS 1005
QY 1007 LMGKNVCPLFTFTRRRFIACPRGGFICPVTGPPSGNRETTLSDOVRGIIYSGGAMVQ 1066
Db 1006 LMGKNACPLLIFTRKFLVACPRAGFVCAASSLGGGAHESSICEQLRGIISGGAAYA 1065
QY 1067 LAIYATVRAVGAQAQMAFDWLSLTDDEFLARDEBELHQIQTLETPTWVEGAL--- 1123
Db 1066 SSVFVATVKSIGRPQTQLOIEDWLALEDEYLSZEMMELTARALRGNGEWSTDALEVA 1125
QY 1124 -EAVKILDEKTAGDGTPTNLAFFND--SCEPSHDTTSNVNLISGNTSGSVPGIKRP 1180
Db 1126 HEABALVSQLNAGE-----VFNDFDQCE-----DDNATPFGGAPGAFAGRKKA 1173
QY 1181 PEDELFDLSGIPKHKHNTWEM 1203
Db 1174 FHGDDPFG-EGPPDKGDLTLD 1195

```

RESULT 6

```

ID DNBI_HSV2H STANDARD; PRT; 1196 AA.
AC P89452;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major DNA-binding protein (infected cell protein 8) (ICP 8 protein).
EN DNP OR UL29 OR ICP8.
OS Herpes simplex virus (type 2 / strain HG52).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
CX NCBI_TaxID=10315;
RN [1]
RP SEQUENCE FROM N.A.
RA Dolan A.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
SC -/- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
REPLICATION.
CC -/- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -/- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; Z86099; CAB06754.1;
CC InterPro; IPR000635; Viral DNA_bind.
CC Pfam; PF00747; Viral DNA_bp; 1
CC DNA-binding; DNA replication; zinc-finger; Nuclear protein.
CC ZN_FING 499 512 C2HC-TYPS.
CC DOMAIN 1168 1196 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

```

```

SQ SEQUENCE 1196 AA; 128423 MW; A19CA843280DD7F5 CRC64;
Query Match 47.9%; Score 3015.5; DB 1; Length 1196;
Best Local Similarity 49.0%; Pred No 1.7e-223;
Matches 595; Conservative 213; Mismatches 366; Indels 41; Gaps 16;
QY 4 TQKTVVTPGPGVYV--ACRVEDLDIEEISFLAARSTDSLALLPLMRLNLTVEKTFSS 61
Db 7 TTTTVKVPFGMGVYVGRACPAEGL--LSLLSARGSDADVAVAPLIVGLTVESGFAN 64
QY 62 LAVVSGARTTGLAGAGITLKTTHSFVPSVFGKXVLPSSAAPNLTRACNAARERFG 121
Db 65 VAAVVGSRITGLGTFVAVSILKMPHSYSPVTVFGRHLAPSTQAPNLTRCEARPHFG 124
QY 122 FSRGCGPVDGAVETTGABICTRLGLEPENTILYLVYVTFKAEAVFMCNLFVHGGDIV 181
Db 125 PADYAPRPCDLKHETTGDCALCERLGLDPRALLYVITEGFREAVCISNTFLHGMKV 184
QY 182 HNHGDIVIRIPLFPVOLFMPPVNRVLPDPNTHRSIGEPVPTPTPYNTGLCHLHDCV 241
Db 185 TIGDAVHRIPIVYPLQMPDPFSRVIAADPPNCNHRISGENFNYPPLPFNRLPARLLFEAV 244
QY 242 IAPMAVALRVNNTAVARGAAHLAFDENHGAFLPDDITVYTFQSSSGTATTARGARRND 301
Db 245 VGPAVALRAENVDAVARAAHLAFDENHGAALPADIITETAESAG--KPORGAR--- 299
QY 302 VNSTKSPSGCFERRILASIMADTALHAENVNTGIYETPTDIDKEWPMFIQWEGTLP 361
Db 300 --DAGNKGPAQGFQRLASVMAGDAALALESIYVMAVDFEPPDDITTWLLEGOETPAAR 357
QY 362 LNALGSVTARVAGVIGAMVSPNSALYLTVEBDSGMTAKDGGGSPENRFYQFAGPHLA 421
Db 358 AGAVGAYLARAAGLVAMVFTNSALHLTEVDGAPADPKDHSK--PSFYRFFLVPGTHVA 416
QY 422 ANQTDREGHVY-----SSQSTGSSNTEFSVDYDALICGFGAPLIARLLFYLERCDAG 474
Db 417 ANQLDREGHVYFGYGRPTAPLVGGTQ--EFAGEHLAMLGFGSPALLAKMLFYLERCDGG 475
QY 475 AFTGGHG--DALKYVTGTFDSEIPCSCEKHTRPVCAHTTVHRLQRMPRFGQATPIGV 533
Db 476 VIVORQEMVFRVAVSGQTDVPCNLCCTFETRIACAHTLMRLARHPKFPASAARIGV 535
QY 534 FGTWNSQYSCDPLGNYPYLIILRKPGDQTEAKATWQDTYRATLERLIDLEQERLLDR 593
Db 536 FGTWNSAYSCDVLGNVYAFSAALKR--ADGSENRTTQMSTYAAATERVMAEALQYVDQ 594
QY 594 GAPCSSGLSVLVDPTRILDLARIEQTTQFMKLVETRDYKIREGLSEATHSM 653
Db 595 AVPTALGRLETTIIGNREALHTVNNIKQLVDREVBOLEMLT:EGENFKFRDGLAEANHAM 654
QY 654 ALTFDPYSGAFCPITNFLVKRTHLAVQDLALSOQCHVYGOQVEGRNFRNQFQVLRRR 713
Db 655 SLSLDPYTCGPCPLQLLARRSNLAVYQDLALSOQCHVYFAGOSVEGRNFRNQFQVLRRR 714
QY 714 FVDLFGGFTSTRITVYVSEG--PVSAPNPTIGQDAPAGRTFGDLARVSVSEVIRDKV 772
Db 715 VMDLFFNGFLSAKTLVALSEGAACIAPSLTAGTAPABSSFGDVARVTIGFPKELAVK 774
QY 773 NRVVFSNCNTLSEAAARLVGLASAYQOEKRVDMHGLGFLKQPHGLLFPFGMPEN 832
Db 775 SRVLFAGASNAASAARVASLQAYQKPKRVILLGLGFLKQPHAVTFPFGKPPG 834
QY 833 SKSPNQFWTLQONQPADKLTHEBITTAAVKRPTEEVAAINFILPTTCIGELAQF 892
Db 835 SNQPNQFWTALQONQPADKLTHEBITTAAVKRPTEEVAAINFILPTTCIGELAQF 894
QY 893 YMANLILKYCHDSOYLINTLSIITGARRPRDPSSVLHWIRKDVTSADIETOAKALLEK 952
Db 895 YMANQILRYCHDSOYLINTLSIITGARRPRDPSSVLHWIRKDVTSADIETOAKALLEK 951
QY 953 TENLPELWTAFTSTHLVRAAMQPMVVLGISIKHGAAGNNRVFOAGNWSGINGCKN 1012
Db 952 LDAHPGAWTSMFASCNLLRPVMAARPMVVLGLSISKYGMAGNDRVFOAGNWSLGGKN 1011

```

1013 VCPFLTFDRTRRRIIACPRGGFCIPVTGPSSGNRRKTTLSQVGRGIIVSGAMVQLAIYAT 1072
 |||||:::||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
 1012 ACPLLIFDRTREVLACPRAGFVCARASSLGGGAHEHSLCEQLRGIIAEGGAAVASSVFVA 1071
 |||||:::||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
 1073 VTVAVGARAHMAFDWDLISLTDDFEFLARDLEEHDQIIQTLEPPTVEGAL----EAVKI 1128
 :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
 1072 TVKSLGPRTQQIQIEDWLALLEDEYLSEEMFEFTRALERGHGWSTDAALEVAHEAEAL 1131
 :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
 1129 LDEKTTTAGGETPTNLAFNDSCEPSHDITSNVLNISSGNSISTVPGLKRPEDEDELFD 1188
 :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
 1132 VSQLGAGE-----VFNFPGFGEDDHASFGGLAAA--AGAAGVARRAFHGDPPFG 1182
 :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
 1189 LSGPIKHGNITWEM 1203
 :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
 1183 -EGPEKK-DLTLDM 1195
 :||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 7
DNEI_HSV2
ID DNEI_HSV2 STANDARD; PRT; 1197 AA.
AC P36384;
YT 01-JUN-1994 (Rel. 29, Created)
MT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
EE Major DNA-binding protein (Infected cell protein 8) (ICP 8 protein).
S DBP OR UL29 OR ICP8.
N Herpes simplex virus (type 2).
C Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
C Alphaherpesvirinae; Simplexvirus.
X NCBI_TaxID=10310;
K [1]
P SEQUENCE FROM N.A.
A MEDLINE=93228441; PubMed=8385914;
X Toh Y., Liu Y., Tanaka S., Mori R.;
T "Nucleotide sequence of the major DNA-binding protein gene of herpes
T simplex virus type 2 and a comparison with the type 1.";
L Arch. Virol. 129:183-195(1993).
C -1- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
C REPLICATION.
C -1- SUBCELLULAR LOCATION: Nuclear (Probable).
C -1- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
C FAMILY.
C PIR; A48350; A48350.
R InterPro: IPRO00635; Viral_DNA_bind.
R Pfam: PF00747; viral_DNA_bp; 1.
R W DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
I ZN_FING 499 512 C2HC-TYPE.
T DOMAIN 1169 1197 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
Q SEQUENCE 1197 AA; 128412 MW; C1576B5B865BF8 CRC64;

Query Match 47.9%; Score 3013; DB 1; Length 1197;
 Best Local Similarity 49.0%; Fred. No. 2.6e-223;
 Matches 595; Conservative 213; Mismatches 367; Indels 40; Gaps 16;

y	4	TQKTVTPTPLGYVY--ACRVEDIDIEEISFLAARSTDSDLLALLPLMRNTVKETTS 61
b	7	TTTTTKVPPGPMGVYGRACPAGLEL--LSLSNRSGDADVAVAPLVIGLTVESGFAN 64
y	62	LAVVSGARTTGLAGATIKLKTTTHFPSPVFFVGXHLVPSSAAPNLTRACNAARERF 121
b	65	VAAVVGSGTTLGLETAVSLKMLPMSHYGPSVTVFVGRRLAPSTCAPNLTRCLRARRHF 124
y	122	FSRCCGPPVDCAVETTGAECITRLGLEPENTILVLVTALPKFAVMCNVFLHVGGLDIV 181
b	125	FSDYAPPCDKLHETTDGALCERGLQDPDRALLYLVITEGFRFAVCISNTFLHUGMDKV 184
y	182	HINHGDIRIPLFPVQFMPPDNRLVDPFNTHRSIGEGVYPTPYNTGLCHLIHDCV 241
c	185	TIGDAEVHRIPVYPLQMFPDPSVIADPFNCNHSRIGENFNYPPLPFNRLARLFEAV 244
/	242	IAPMAVALRVNRYTAVARGAHLAFDENHGAVLPPDITVTFYFOSSSGTGTARGARN 301

QY	358	TLPRNALG	SYTARVAGVIGAMVSPNSALYLTVE	DSGMTAKDGGPSPNRFVQAG	411			
DB	350	TAARAASIG	AYLGRAAGVIGAMVSSNSALHLTE	VDGADPADPKD-PTKPSYRFLVPG	408			
QY	418	PHLAANPQ	TRDRCHVLSQS-----TG	SNTFSVDYLALICGFCAPILARLLFV	471			
DB	409	TYVAANPQ	LDRGRVAGHEG	RPVPIVGGNHEFTCEHLATLCGSP	ELAKWLYLBERC	468		
QY	472	DAGAF	TG-CHGDALYKVTGTDS	EIPCSLCEKHTRPVCANTVHRLR	MRPFGQATRP	530		
DB	469	DGCVILG	PEMDTFKYVSDSAHTD	VPCLCSLDRNRSCAHTILLR	ASHPKFTTIRGA	528		
QY	531	IGVFGTMS	SOXSDCDPLGNAYPILILK	PKGDQTEAAKATMODTYRATLE	FLDLEQEL	590		
DB	529	IGFGVMS	AYSDDCVLGNVASF	AIKRM-DVQETARALM	QETYSRVSERVVAELENLNY	587		
QY	591	LDRGAPCS	SEGSSVVDVHPTRIL	TLRARIQTTFQPMKVL	YETRDYKIREGLSEAT	650		
DB	588	IDAAVPT	SPAKLEIITGREALQ	TVSNVQVVDGVAQLM	ALVEGRGRPREALGEAN	64		
QY	651	HGMALT	FPYSGAF	PIPLFKVTHLAVQDIAL	SOCHCVYGOQVEGRFRAQF	PVL	710	
DB	648	HANSITL	PHASVCP	LLOMLGRSNLAVQDIAL	SOCHGVFEQAV	GRNFRSQFQV	707	
QY	711	RRFV	DLFNGGFI	STRSITVLTSEGP-VS	APNPTLQDAPAGRTD	GDGLARVSVEVIRDI	769	
DB	708	RRRV	LDMFNNGFL	SARTILTVALTGACIS	APGLVSGHAAESSG	EGDVARVNLGFPKEI	767	
QY	770	RVKR	RVVFSG	CTNLSEAPARLVGLAS	AYORQEKVDMHCGAL	GLFKQFHGLLPFRGM	829	
DB	768	RKSV	LVFAGAG	PAASEAPARLAG	QSAVQSDKRDVILLG	FLGFLMUKQFHATL	FPNGK	827
QY	830	PNSKS	PNPQFW	TLORNQPADKLTHEE	ITIAAKRFTTEYAAIN	FINLPPTCIGEL	889	
DB	828	PFGS	DNPNQFW	TALORNQPARLL	SREDISLIAFVKFES	VEYCGAGNFVLPNNISEL	887	
QY	890	ACFYMAN	LILKYCDHSOYL	INTLTSITCARPRDP	SSVLHWIRKDVTS	SAADIETOAKAL	949	
DB	888	AMYMAN	QILKYCDHSY	TFINTUTALIGSR	PPNQAQAAAWAP	PGGPT---ELEAQARS	944	
QY	950	LEKTEN	PELWTAFT	STHLVRAAMN	QRMVVLGISISKYHGA	AGNNRVFOAGNWSGLNG	1009	
DB	945	VANP	GDHPGAWT	TFASONLLRPVAT	REMVVLGISISKYTG	MAGNDRVFOAGNLANLIG	1004	
QY	1010	GKW	CPLFTDRTER	FIACPGGFC	VPVTPGSSN	RETTYLSQVRGII	VSGAMVOLAI	1069
DB	1005	GKNAC	PLIITDRTRK	VFVACPRAGF	VCAVSAVSGAH	ESSICEQRAIIA	IEAGGATVASDV	1064
QY	1070	YATV	RVAGARACHMA	FDWLSITDDEFLAR	DLERLHDOIITOLET	PWTVEGALEAVKIL	1129	
DB	1065	FAAAK	SILGARVQOLQ	IEDWALLLED	EVISEMMELAGRAL	ERGGENSDAALD	AVAREA	1124
QY	1130	DEKT	-----AGD	GETPTNLAFN	FD-----CEPS	HDTTSN	1160	
DB	1125	EMVTR	HVDAEETP	DGFAED	GDGADAGLAVH	LSRRPLACSLD	FGDAPAEKENDITLD	1184
QY	1161	ML	1162					
DB	1185	ML	1186					

RESULT 9

DNBI	HSVE1	STANDARD;	PRT;	375 AA.
ID	DNBI HSVE1			
AC	Q03444;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Major DNA-binding protein (fragment).			
GN	31.			

OS Equine herpesvirus type 1 (isolate HVS25A) (EHV-1).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.

CC Alphaherpesvirinae; Varicellovirus.
 CC NCBI_TaxID=10327;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=94106109; PubMed=8279122;
 CC Bell C.W., Whalley J.M.;
 CC "Herpesvirus ICP18.5 and DNA-binding protein genes are conserved in
 CC equine herpesvirus-1";
 CC Virus Genes 7:219-228(1993).
 CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
 CC REPLICATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC R EMBL; D13930; BAA03033.1; .
 CC R PIR; J00846; J00846.
 CC R InterPro; IPR000635; Viral DNA bind.
 CC R Pfam; PF00747; Viral DNA bp; 1.
 CC R DNA-binding; DNA replication; Nuclear protein.
 CC R NON-TER 1
 CC 2 SEQUENCE 375 AA; 40309 MW; ECF327925EBE999B CRC64;
 CC
 CC Query Match 17.4%; Score 1094.5; DB 1; Length 375;
 CC Best Local Similarity 55.5%; Pred. No. 1.9e-76;
 CC Matches 212; Conservative 57; Mismatches 98; Indels 15; Gaps 5;
 CC
 CC 828 GMPNKSPPQFWTLQNNQMPADLTHEITIAAVKRFTTEVAAINFILPTCIG 887
 CC 2 GHPPGIDTPNQFWTLQNNQMPARLLSKEDTETIAKRFDSYSAINFILPTNNG 61
 CC
 CC 888 ELAQFYANILKYCHDSQVLYINTLSITGARRPPDSVLHWIRKDVTSAAIDTQAK 947
 CC 62 ELAQFYANILVLYCHDSQVLYINGLTAVVGSRRPRDPAALAVLAWIDRTINGAADVPAQA 121
 CC
 CC 948 ALAEKTENLPETLTAFTSTHLVRAANNQPMVVLGISIKYHGAAGNNRVQAGNWSGL 1007
 CC 122 EVLQRLGNSNAATGFTSTNNWRYVMDQPMVVLGISIKYSGAGNNRVQAGNWSGL 181
 CC
 CC 1008 NGKKNVCLPFTFDRTRFIACPRGGFICPVTPGSSGNRETTLSQVRIIVSGGAMVOL 1067
 CC 182 NGKKNVCLPMAFDTRFRFLACPRVGTCEAGGFGVRENTLSEQVRGIVSEGGPMVQT 241
 CC
 CC 1068 AIYATVVRVGAQAQMAFDNLSTDDFLARDLELDHDIQIOTLETPTWVEGALEAVK 1127
 CC 242 AVFAAVLHALGATQHLAVDDMTGLVDDBFLAASLDALNATVVDQ-GEWSVEAAQELVK 300
 CC
 CC 1128 ILDEKTTAG---DGETPTNLAFNFDSC---ERSHDTTNSVLNLSGNSITSGSVPLGKRP 1181
 CC 301 NMEAQTITAGAVAGEG---AFDFGACVGTFTQOSTSA-----FNGGLMAAAPAGKRS 352
 CC
 CC 1182 EDDELFDLSGIPKHNINTEM 1203
 CC 353 PDDILDMGAPPEKSGGLTFDM 374

SULT 10

BI_EBV

CNEI_EBV

STANDARD;

PRT; 1128 AA.

P03227;

21-JUL-1986 (Rel. 01, Created)

21-JUL-1986 (Rel. 01, Last sequence update)

16-OCT-2001 (Rel. 40, Last annotation update)

Major DNA-binding protein.

BALF2.

BALF2.

BALF2.

BALF2.

OS Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 CC Gammaherpesvirinae; Lymphocryptovirus.
 CC NCBI_TaxID=10377;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=84270667; PubMed=6087149;
 CC Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 CC Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
 CC Tuffnell P.S., Barrett B.G.;
 CC "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 CC Nature 310:207-211(1984).
 CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
 CC REPLICATION.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; V01555; CAA24808.1; .
 CC DR PIR; A43045; Q0BE47. Viral DNA bind.
 CC DR InterPro; IPR000635; Viral DNA bp; 1.
 CC DR Pfam; PF00747; Viral DNA binding; Zinc-finger; Nuclear protein;
 CC DNA-binding; DNA replication; Zinc-finger; Nuclear protein;
 CC Early protein.
 CC KW Early protein.
 CC SQ SEQUENCE 1128 AA; 123121 MW; 283D5258031A16D6 CRC64;
 CC

Query Match 10.6%; Score 669; DB 1; Length 1128;
 Best Local Similarity 23.2%; Pred. No. 5.8e-43;
 Matches 292; Conservative 186; Mismatches 537; Indels 246; Gaps 45;
 QY 13 GPLGVACRVDELLEISFLAARSTDSLALLPLNLTVEKTFSSLVAVSGARTTG 72
 DB 18 GPCGYIFYPYPLATVLEVAVLGTGYAGHRCVLLCGITVPGFSIN---VKALHRRP 74
 QY 73 LAGAGITLKLTTSHFYFVFVFGHGKVLPSAANPLTRACNAARERFGFS-----RCQ 126
 DB 75 DPNCGL---LRATSYHRDIVVFHNAHWVPPIFEGPGLEALCGETREVFYDAYSAIPRS 131
 QY 127 GPPVDGAVETGAEICTRLGLEPENTILYLVLTALFKAVFMCNVFLHYGLDVIHNG 186
 DB 132 SKPGDFFPE-----GLDPSAVLGAVAITFAFKERLYSGNLVAIPSLKQEVAVGQS 181
 QY 187 DVIRIPLPVQLFMPDVNRLVPPDFNTHRSIGEGFYPTFPYNTGLCHLHDCVIAPIA 246
 DB 182 ASVAVPLDYKEVPEGPVQL-----RQFVNSDLSCRMHEALVTGLA 222
 QY 247 VALRVNVTAVAGAHAFLAFDENHEGAVLPDITYTYFQSSSGTITARGARNVDNSTS 306
 DB 223 QALKRVREVGKL-----VELLEKQSLQDQAKVAVKAPLKEPPAST 261
 QY 307 KPSPSGGFERRLASIAADT-ALHAEVIFNTGYE---ETPTDIK---EWPMTFIGEGTLP 360
 DB 262 ISHPDSG-----ALMIVDSACELAVSVAPMLASHEHETPASLNYDSWPLFADCEGPEA 315
 QY 361 RLNALGSYARVAGVIGAMVFPNSALYLTVEDESGMTEAKDGGPGSPFNRFYQAG--- 417
 DB 316 RVAALHRYNASLAPHVSTQIFATNSVLYV-----SGVSKSTGCKESLFSNRYHGLGT 370
 QY 418 -----PHLAANPQTDRCGHVLSQSTGSSNTEFSDYLYLALICGAPILARLLF 466
 DB 371 LQSGTWDPCKRCPFCGWMGPDVTG-----TNGPGN--YAVEHLVYAAFSFSPNLLARYAY 422
 QY 467 YLBRCDAGFTGGHGDAL-----KYVTGTFTDSELPSCLEKHTRPVCAHTTVRLRQR 519
 DB 423 YLQFCQ-----GQKSSLTFVPETGTSYVAGAAASPM-CSLCEGRAPAVCLNTLFFRLRDR 475

```

QY 520 MPREGATQPIGVGTGTMNSOYSCDPLGNVAPYLIL-----RKPQDTEAAKATMQDYTR 575
DB 476 PPPVMSQTRDPYVYISAGSGYNTEDFLGNFLNFDKEDQQRDPDEPRYTYWQLNQNL 535
QY 576 ATLRLFLIDLEQERLLDRGAFPCSGSEGSLVIVDHTFRRILDLTLRARIQOTTTOFMKVLV 635
DB 536 ERLSRLGIDAEGLEKEPHGP-----RDFVKMFQDVAADVAEVQFMNSWA 582
QY 636 ETRDYKIREGLSEATHSMALTFDPSGAFPCINFLVKTTHLAVQDLALSCQCHVFYQG 695
DB 583 K-NNITYKDLVKSCYHVQNSCFNFAQPCFIPTQLFYRLSLTILQDISLPIQWC----- 636
QY 696 QVEGRNFRNQFVLRFRFVDFLNGGF-----ISRSITV-----TLSEGVSAFNPGLG- 745
DB 637 -----YEND-NPGLGOSPENLKGHTCTLTFWRSIAIDKGVLTAKAVHGEPTCOL 689
QY 746 -----QDAPAGRTFGDLARVSVEVIRDIRVKNRVVFGNCTNISEARARLVGLASA 798
DB 690 PDLDAALQGRVYGRRLPRVRSKVLMLCPRIKIKRVVFTGENAALQNS-----FIKS 742
QY 799 YQREKRVMDLHGALGELLKQFGLLFPFGMPNPKSPNPQWFWTLQRNQMPADK-LTH 857
DB 743 TTRRNVY--INGPYMKFLNTHKTLF-----PDTKLSSLYLWHNFSRRSRVPVPGASA 795
QY 858 EBITTAA-VKRTTEEYAAFINLPPTTCIGELAQFYMANLILKYCDHSQYLNTILTSII 916
DB 796 EYSDLALFVDGGSRAHEESNVIDVVGNLVTYAKQRLNAILKACGQTFYISLIQGLV 855
QY 917 --TGARRPRDPSVLHWRKDVTSADIEIQAALKEKTENIPELWTTAFTTHLVRAAM 974
DB 856 PRQSVAPADYPHVLG--TRAVESAA-----AYAEATSSLTAT-TVVCAATDCLSQVC 905
QY 975 NORPMVVLGISISKYHGAAGNNRVFOAGNWSGLNGKVCPLTFEDTRRFTIACPRGG- 1033
DB 906 KARPVVTLPTVINKYTGNGNNOIFQAGN-LGVFMGRGV-----DRN---LLQAPGAGL 955
QY 1034 -----FCPTVGPSGNR-ETTLSDQVGRGIIVSGGAWV-----QLAIYATV 1073
DB 956 RKAQSGSSMRKKEVFVATPTLGTUVRKRTQAATYEIENIRAGLEAIIISQKQBEDCVFDV 1015
QY 1074 VRVAGARAQHMADDDWLSLT--DDEFLARDLEELHDOIQTLET-----PWTVEGALEA 1125
DB 1016 CNLVDAMGACA-----SLTRDAEVLGRFVLDVSLVLETIATASSGIEWTAAANDF 1070
QY 1126 VKILDEKTTAGGETTNIAFNFDSCPESHDTNVLNLSGNSIG-----STVPGL 1177
DB 1071 L-----EGVWGGPAAQDNFISVAEPVSTASQASAGLLGGGGGGGRRKRRLATVLPGL 1126
QY 1178 K 1178
DB 1127 E 1127

```

RESULT 11

DNBI_HSVSA

ID DNBI_HSVSA

AC P24910;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Major DNA-binding protein.

EN 6 OR KFRF1.

SC Herpesvirus saimiri (strain 11).

SC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

SC Gamaherpesvirinae; Rhadinovirus.

SC NCBI_TaxID=10383;

UN [1]

IP SEQUENCE FROM N.A.

IP MEDLINE=90163221; PubMed=2154888;

IP Albrecht J.-C., Fleckenstein B.;

IP "Structural organization of the conserved gene block of Herpesvirus

IP saimiri coding for DNA polymerase, Glycoprotein B, and major DNA

binding protein.";

Virology 174:533-542 (1990).

[2]

RP SEQUENCE FROM N.A.

RX MEDLINE=93333688; PubMed=1321287;

RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,

RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,

RA Honess R.W.;

RT "Primary structure of the herpesvirus saimiri genome.";

RL J. Virol. 66:5047-5058 (1992).

[3]

RP SEQUENCE OF 1-81 FROM N.A.

RX MEDLINE=91205588; PubMed=1850023;

RA Nicholas J., Coles L.S., Newman C., Honess R.W.;

RT "Regulation of the herpesvirus saimiri (HVS) delayed-early 110-

RT kilodalton promoter by HVS immediate-early gene products and a

RT homolog of the Epstein-Barr virus R trans activator.";

RL J. Virol. 65:2457-2466 (1991).

CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA

CC REPLICATION.

CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN

CC FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC

CC EMBL; X64346; CAA45629.1; -

CC EMBL; M31122; AAA46162.1; -

CC EMBL; M60849; AAA46157.1; -

CC InterPro; IPR000635; Viral DNA bind.

CC Pfam; PF00747; Viral DNA bp; 1.

CC DNA-binding; DNA replication; Zinc-finger; Nuclear protein.

CC ZN_FING 453 466

CC ZN_FING 453 466

CC SEQUENCE 1128 AA; 127457 MW; BA3ADE861381328D CRC64;

CC

Query Match 9.6%; Score 604; DB 1; Length 1128;

Best Local Similarity 22.9%; Pred. No. 5.9e-38;

Matches 275; Conservative 189; Mismatches 509; Indels 226; Gaps 40;

QY 11 PTGPIGVACRVEDLDLEISFLAARSTDSLALPLMRNLTKVTFTSSLAIVSGART 70

DB 18 PVFPCGIIYYPKEGPPFKESALLGNKNVVGAMSLPLSLDITVESNFSFNKAVHKXID 77

QY 71 TGLAGAGITKLTTSHFYPSVFVFGKHVLPSSAAPNLTRACNAARERFGRSQGPPV 130

DB 78 M-----TLLVRVSAYHREAIVFNTDLFEIFVGPGLDILCSARSILFYTNF--VPR 129

QY 131 DGAVETTGAEICTRIGLEPENILVLTALKEAVFMCN-VFLHYGGLDIVIHNGDVI 189

DB 130 TDLRDTVDIKOLYAPPYSDSCFMAVWVTEGKERLYFGNLVPIIAQGLK-VQINGREAV 188

QY 190 RIPLFPVQLFMPDVNRLVPDPENTHRSIGEGVYPTPTFNTGLCHLHDCVIAPMVAL 249

DB 189 KIPLXDEDLFSKSHL-----PRFYTPSVSKYLHDSVFTSIAQAL 229

QY 250 RVNVTAVARGAAHLAFDENHEGAVLPDITVYFQSSSGTTTARGARNVDNSTKPS 309

DB 230 RIRDVESVIRASEKSIQOYKLAKI-----VNSK----- 259

QY 310 PSQGFERRILASIMAADTA-----LHAEIFNTGI-YEETP-----TDIKEWPMFIGM 355

DB 260 -----DFSLSQVKCODASAFWIDICIAELAMSICYLSFLEAPQDCAVLDTYSWIFETA 314

QY 356 EGTLPRLNALGYSYARVAGVIGAMVFSNLSALYTEVEDSGMTEAKGGPGPSFNRYQF 415

DB 315 ETEEDRIKAIDQWNAAMSVHYVTHLFTNSVLYLTINKQTSKNKSEQN---VYNTYFMQ 371

```
2Y 416 AGHLLANPQTRDGHVLSQSSTNTSFEVDYLALICGFGAPLLARLLFYLERCDAGA 475
b 372 HGLVAADATQRENGEPAPSGAPKSGGGTYLHLALASSFSFHLARNCYMQPCQHQK 431
2Y 476 FTGGHCDALKYVTGFDSEIPSLCEKHTRPVCAHTVHLRQRWPR-FGQATROPIGVF 534
b 432 STINANYSVQVGTAAASDLCELCQGTCPASCINTFLYLRKDRPPVGLSGQRDPYVVT 491
Y 535 GTMNSQYSCDPLGNVAPYLIIRKPGDQTEAAKATMQDTYRATLERLFDLEQERLLDRG 594
b 492 G-VSQYNDLWLGNFATF--REKEDRAQNAESEKTYWQLQNVV-----EKLSTWG 542
Y 595 APCSEGLSSVIVDPTTRRLDITURAEIQTTQFMKVLVETRDYKIREGLSEATHSMA 654
b 543 VTEGTVG-SELITDIQSELKTPRDINDVVDSEVKFMCLVK-NNINPREIKTVHVLH 600
Y 655 LTDFDYSAGFCPTNFLVKRTHLAVVQDLAL-----SQCHCVFYQQOV 697
b 601 YCCNVFWQAPCMFNLFYKSVLAIQIDICPIAMTYEQDNPS:GMPSEWLKVHY--QT 658
Y 698 EGRNPR-----NQOPVLRRRFVDLFGGFISTRSITVTLSGPGVSAENPTLGO 746
b 659 IWTNFKSSCLDRGLVTGSEHKIVHTDMFCDELN-----IDSALS-----GQ 699
Y 747 DAPAGRTDGLARVSVEVIRDIRKVRVFSNGCTNLSEARARLVGLASAYORQEKRV 806
b 700 IVP--MKMQVRLAKALLTPKTIKKNRIVFSN--SSMTETIQSGFIKSAT-----KKDS 750
Y 807 DMLHGALGFLKQFHLGPRGMPNPSKSNPQFWFLLQNMPPA-DKLTHEBITTIA- 864
b 751 YIVTPYKFLNSLHKWF-----PNAKISALYUHTFSQKKQLPVLPGISRENWELAN 805
Y 865 AVKPTTEYAAINFINPPTCIGELAPYMANILKYCDHSQYLINTLSI-----ITG 918
b 806 YVETSSKMDHMDVLDIPTLLYVAKVRLNNTILRTCGGTQFVATTLOCLLPTLOTISA 865
Y 919 ARRRP---DPS--SVLHWIR--KD-----VISAADIETQAKALEKTENBELMTTAF 964
b 866 TEYPHVLLDQSIMVDHYLSIKDKHALTVQTTUKEDIATVGK----- 908
Y 965 TSTHLVRAAMQRMVVLGISISKYHGAAGNNRVFQAGN---WSGLNGKNVCPLFT-PD 1020
b 909 -----QRPIVTVPLVANKYTCINGNTQIFQCNGLGYFNGRVDENLIPDSTGR 957
Y 1021 RTRFLIACRGGFICP-----VTGPPSGNRET-----TLDQVRGIIVSGAMVQLAIVA 1071
b 958 RQNNSSWRRHRHVFMTPMVAHLVKCNLNLLTFEVEITKKNVQNFEDKD---NLNIFD 1014
Y 1072 TV-----VRVAGARQAQMAFDLWLSLTDDEFLARDLELHDLQIITLETPTVEGALEAV 1126
b 1015 NVVLELVKGLGDSCEINTEDDLQFLGYEYVIMSDEIWSRFQIITDSGAPMSVENVTKVL 1073
RESULT 12
DBI HSV6U STANDARD; PRT; 1132 AA.
> P52338;
> 01-OCT-1996 (Rel. 34, Created)
> 01-OCT-1996 (Rel. 34, Last sequence update)
> 16-OCT-2001 (Rel. 40, Last annotation update)
> Major DNA-binding protein (MDBP).
> U41.
> Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).
> Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
> Betaherpesvirinae; Roseolovirus.
> NCBI_TaxID=10370;
> [1]
> SEQUENCE FROM N.A.
> MEDLINE=95266321; PubMed=7747482;
> Gompels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
> Martin M.E., Efstathiou S., Craxton M., Macaulay H.A.;
> "The DNA sequence of human herpesvirus-6: structure, coding content,
> and genome evolution."
```

Virology 209:29-51(1995).

-!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA REPLICATION.

-!- SUBCELLULAR LOCATION: Nuclear (Probable).

-!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN FAMILY.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X92436; CAAG3167.1; -

EMBL; X83413; CAA58375.1; -

InterPro; IPR000635; Viral DNA bind.

Pfam; PF00747; viral DNA bp, 1

KW DNA-binding; DNA replication; Zinc-finger; Nuclear protein.

FT ZN FING 459 475

SQ SEQUENCE 1132 AA; 127762 MW; 4468D3B559F02D8 CRC64;

Query Match

Best Local Similarity 7.8%; Score 492.5; DB 1; Length 1132;

Matches 274; Conservative 185; Mismatches 544; Indels 253; Gaps 53;

QY 2 ENTQKTVPTGPGLYVYACRVEDLDLEEISFLAARSTDSDLALLPLMRNLTVKRTFS 61

Db 3 DENETWSAPVSTAAMIYVFPKRELLDVLVSLMERNSPIVISPLMLNLTVDNFSST 62

QY 62 LAWSGARTTGLAGAGITLKLTTSHFYSVEFVGKHLV-PSAAANLTFACNAARERF 120

Db 63 V-----KPTITPFGGILTKITSFMPVCFPGTQOLVGMAEDHGLRLCEQTKF 115

QY 121 GFSRCQGPVPGAVETTGAEICTRLGLEPENTILYLVVLTALFKEAVFMCNVLHYGLDI 180

Db 116 HLOSPEVPTARKVIDIKA--LCSAVGKDADSVICHVACGNGFKELLF-----AGLI 165

QY 181 -----VHNHGQVIRIPLFPVQLFMPDVNRLVPPFNTHRSIGEGFYPTFFVTG 232

Db 166 PCVEBQIOVQGEVSCVKIPLYSATLFETE-----ETISLSSCTEFIOBERGFLPA 216

QY 233 LCHLHDCVIAFMAVALRVNRVNTAVARGAAHLAPDENHEGAVLPDITYVFOSSSGTT 292

Db 217 LSETLFYVYFTSWGTTLRFPNTKELIDAGLKQFDQGEQIVKLAPHKTYL----- 266

QY 293 TARGARNVNSTKSPSGGFERRLASIMAADTALHARVFN-----TGIYEETPTDI 346

Db 267 ---GISGQKISAVEKD-----FLMLVDSV--TELSFSGHVAEYLDVVD--PSQI 309

QY 347 ---KEWPMFIGMEGTLPRNLALGSYARVAGVIGAMVFPNSALYLTEVE-----D 394

Db 310 MNFNDWPIIRNSETHAERMAQLNKLHLSHSLAVLIFAPNSILYCKLAFIPNVKQAFN 369

QY 395 SGMTAKDGGPGSFNRFYQFAGFHLAANPQTRDGH-VLSSQSTGSSNTEFSDVYDALI 453

Db 370 SVMTQ-----ELLRSLSFCNALSSDDVYNDNRKLIKCDSTSGKDDKFSANHLA 422

QY 454 CGFGAPILARLLFYLERCDAGATGGHGDALKVTVTTFDSEIPCSLCEKHTRPVCAHTV 513

Db 423 CATSPQLLSFVWNLNRMSVYNAGNAHTEIYNHLVNC--SALNCEFCGKCCQSCIGTAM 480

QY 514 HRLRQRFREFQATROPIGVFTMNSQYSCDPLGNVAPYLIIRKPGDQTEAAKATMQDT 573

Db 481 VRVGTRLPAIPKVKKEPLVMSMFSEYAAVVDILGSFG-----RKFSVSELKIGKQQNT 535

QY 574 YRATLERLFDLEQERLLDRGA-----PCSEGLSSSVIVDHPTR-----RILPT 618

Db 536 LS-----LDRKEVSQLFDYCKNSLIDPVTGDTFNVRSKKDFVSIING 580

QY 619 LRARIEQTTQFMKVLVETRDYKI--REGLSEATHSNALTFDPSGAFCTITFLVKRTHL 677

581 LTQCEECVS---RCIVEMRTQTFRQIENCLOSFNVDTPYATASPFLTSTYKVL 637
DR PIR; T44001; T44001.
DR InterPro; IPR000635; Viral DNA bind.
DR Pfam; PF00747; viral DNA bp; 1
KW DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
FT ZN_FING 459 475 C4-TYPE.
SQ SEQUENCE 1132 AA; 127817 MW; B39AB4E6A5103EFF CRC64;
Query Match 7.8%; Score 490.5; DB 1; Length 1132;
Best Local Similarity 22.0%; Pred. No. 3, 2e-29;
Matches 278; Conservative 180; Mismatches 535; Indels 273; Gaps 54;
QY 2 ENTQKTVTVTPGPGYVYACKVEDLLEISFIAARSTDSDALLPLMRLNLTVEKFTSS 61
DB 3 DENETVSAPYSTAAWYVFPKKEKELDLVLSVLSMERNSPVVISPLMLLTVENDFSTT 62
QY 62 LAVVSGARTTCLGAGITLKLTTSHFVPSVVFHGGKVL- PSSAAPNLTRACNAARFP 120
DB 53 V-----KTPITNFGTITLKITSPMPVCFPHGTETQLVGMADHGLRLCEQTRQKF 115
QY 121 GFSRCQPPVDGAVETTGAGICTRLGLEPENTILYLVVTFALFKEAVFMCNVLHYGGDI 180
DB 116 HLQSFVEPTARKVIDIKA--LCSAVGKADSVICHVACNGGFKELF-----AGLLI 165
QY 181 -----VHLNHGDVIRIPLPVQLEFMPDVRNLYPDPFNTHRS-GEGRVYTPPNTG 232
DB 166 PCVEEQIQVQGEYSCKIPLYSATLFETE-----ETLSLSCSTEFQBERGFPLA 216
QY 233 LCHLHDCVIAPIVALRVNVTAVARGAAHLAFDENHEGAVLPDITITYFQSSSGST 292
DB 217 LSETLFYVVTSMGTLTFRFNTKELIDAGLKQFQDGEQTVKLAPHKYL----- 266
QY 293 TARGARNDVNSKSPSPGPFERRIASIAVATUHAHVFN-----TGIYEPTPTDI 346
DB 267 ---GISGOKISAVEKD-----FLMLVDSVV-TLSFSHVAEYLDSDVYD--PSQI 309
QY 347 ---KEWPMFIMGECTPLRNALGYSYARVAGVIGAMVSPNSALYLTEVE-----D 394
DB 310 MVENDFIIRNETHAERMAQLTNLKLHSSHLAVLIPANSILYCSKLAPIENVQAFN 369
QY 395 SGWTEAKDGGPSPFNRYOFAGFHLAANPQTRDGH-VLSQSTGSSNTEFSVDYLA 453
DB 370 SVMTQ-----ELLRLSLSFCNALSLTSDVYDNRKIKCDSTSGDKKFSANHLA 422
QY 454 CGFGAPILALLFLERCDAGAFGGHGDALKYVTCFDSFI-----PCSLCEKH 503
DB 423 CATCPQLLSVVMNLRMSV-----YNAGNAYTEIYNHLVNSANLCEFCOK 470
QY 504 TRPVCAHTTVHRLQRMFRFGQATROPIGVGTMSQYSDCDPLGNAPYLYLIRKPGDQ 563
DB 471 CCOSCIGTAMRVVGTLPALPKNVKPEPLVMSFYSYAEVDILQSGF-----RKEVSEL 525
QY 564 EAAKATMODTYRATLERFLDLEQERLLDGA-----PCSEGLSSVIVDHPFTR--- 613
DB 526 KEIKGQOQNTLS-----LDRGFVSQIFDYCKNSLIDPVTGDFNVRS 570
QY 614 -----RILDTLRARIEQTTFQFMKVLVETRDYKI-REGLSEATHSMALTDFDYS 667
DB 571 KDFVSIHLGTOCIECVS---RCIVEMRTQTFRQIENCLOSFNVDTPYATASP 627
QY 668 TNFLVKTTHLAVVQDIALSOCHCVFVGQVEGRNFRNQFVLRFRFVDLF-----N 719
DB 628 LTFAYKVLTVLQNLAL-----IVASGHVVDPRCTGNSISKWLQVQYQSLYGT 683
QY 720 GGFSTRSIVTVLSEGPAPNPTLQO--DA---PAGR-----TFDGLARVSVEVRDIR 770
DB 684 KGFNLTFTVKV-----ASNVDMEQILDCLYSGKVKYKVTIOAKCLRSNQCULDR 735
QY 771 VKRVVFGNCTNLSAARLVGLASAYQREKRVDMHGLGFLKQFHLGFLPFRGMP 830
DB 736 IKRPFNKSKTAHNNPYKKNV-----KHKKNPLSGCISLFLPKYHDKLF----- 781
QY 831 PMSK-SPNPQWFVTLQRMQPADKLTHEEIT-IAAVKRFTEEYAAINFLPTCIGE 888

581 LTQCEECVS---RCIVEMRTQTFRQIENCLOSFNVDTPYATASPFLTSTYKVL 637
678 AVVQDLALSOCHCVFYQGOVEGRNFRNQFVLRFRFVDLF-----NGGFSTRSIT 729
638 TVLQNLAL-----IVASGHVVDPRCTGNSISKWLQVQYQSLYGTFFHSSLYKKGFLNTRTVK 693
730 VTLSEGPVAPNPTLQO--DA---PAGR-----TFDGLARVSVEVRDIRVKNRVFSGN 780
694 V-----ASNVDMEQILDCLYSGKVKYKVTIOAKCLRSNQCULDRFRIKRPNKSS 745
781 CTNISEAARLVGLASAYQREKRVDMHGLGFLKQFHLGFLPFRGMPMSK-SPNPQ 839
746 KTAHNNPYKKNV-----KHKKNPLSGCISLFLPKYHDKLF-----PNVKISLLEL 791
840 WFWTLQRMQPADKLTHEEIT-IAAVKRFTEEYAAINFLPTCIGELAQFYANLI 898
792 WQRFLLNNVKTLDIGNPEEVKTFIKAFSITNTYDEIDIDIQECLSTFDICYFHNKF 851
899 LKYCHSQYL--INTLTS-IITG-----ARRPRDS--SVLHWIKDVTSAADTET 944
852 LSALGFHDYLTSLHGLTSKLVTONPVLFPVVDLQKPFSSIQEYLYVYKVLVDG--VEN 909
945 QAKALLEKTENLPETWTATSTHLVRAAMNQPMVVLGISISKYHGAAGNNRVFOAG-- 1002
910 PVIASLSKEPN---FQTIPTS-----RSLVTFGLTEKFPVSLA-NREYFQFQL 954
1003 NWSGLNG-GKNVCPL-----FTFDRTRFIIACPRGGRICPVTGSSGNRETTLSQD-- 1053
955 GWIGSGVDNRLNPTTSALQDFMR-OXTIIIAIKFSEVIV-----KVVRRBAINFEDTEV 1008
1054 VRGIIVS-----GGAMVQIAIYATVVRVAGAAHMAFDMLSLTD--EFLARD----L 1102
1009 VKGKVLSEVENLTNDIDPELLIAEVMR-----DREDKPTMDMLFFVDGREALAASIMKL 1065
1103 EELHDOIQIOTLEPTWTEGALEAV-----KILDEKTAGD-----ETPT 1142
1066 NHLVDVNVKDFSI-TLQSVFVSSNDAPVDFSEILAEEDDQGNLKCDETET 1120
RESULT 13
ID DBI HSV62 STANDARD; PRT; 1132 AA.
AC P52538;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major DNA-binding protein (MDBP).
EN U41 OR XA2L.
SS Human herpesvirus (type 6 / strain 229) (HHV6).
SC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
SC Betaherpesvirinae; Roseolovirus.
XX NCBI_TaxID=36351;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=95074921; PubMed=7983761;
RX Stamey F.R., Dominguez G., Black J.B., Danbaugh T.R., Pellett P.E.;
RT "Intragenomic linear amplification of human herpesvirus 6B orilyt
RT suggests acquisition of orilyt by transposition.";
RL J. Virol. 69:589-596(1995).
CC -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
CC REPLICATION.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC

```
Db 782 PNVKISLELMQRFLLNNVKTLDIGNPEVVKFIKFAPIITNTYDEIDIDIQPECLST 841
Qy 889 LAQYMANLILKYCDHESYL--INTLTS-IITG-----ARRRDPSS--SVLHWIRK 934
Db 842 FIDCFYFNKLSALGFHDYLTSLHGLTSKLVTONVLPVFLVDKQPKFSSIQBYLYVVK 901
Qy 935 DVTSADIETQAKALEKTENPELMTAFTSTHLVRAANNQRPVVLGISISKYHGAAG 994
Db 902 LVLDDG--VNPVVIASLSKEPN---FOTIETS-----RSLVTFGLTKLEFVSLA- 944
Qy 995 NNRVFOAG--NWSLNG-GRNVCP-----FPTDTRFIACPRGGFICPVTPGSSGN 1045
Db 945 NREYFGQGLWIGGSGVDNRNPTLSALQDFRMR-QKTIITATKPESEVIV-----KKVR 998
2y 1046 RETLSDQ--VRGIIVS-----GGAMVQIAIVATVVRVAVGARAQHWAFDDMLSLTDD-E 1096
Db 999 RETIMFDEVVVKVLSIVENLNDIDPELLIIAEVWR---DREDKPTWDDMLFFVDGR 1055
2y 1097 FLARD-----LEELHDQIIQTLFPTWVEGALEAV-----KILDEKTTAGDG--- 1138
Db 1056 ALAASIMLNHLNLMVNRVDFSA-NLQSVFEAVSSNDAPVYDFSEILAEEDDQASGVLK 1114
2y 1139 --ETPT 1142
Db 1115 CDETET 1120

RESULT 14
MBI_HSV7J
D DNBI_HSV7J STANDARD; PRT; 1131 AA.
T 01-OCT-1996 (Rel. 34, Created)
T 01-OCT-1996 (Rel. 34, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Major DNA-binding protein (MDBP).
IN U41.
S Human herpesvirus (type 7 / strain J1) (HHV7).
X Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
X Alphaherpesvirinae; Simplexvirus.
X NCBI_Taxid=57278;
X [1]
X SEQUENCE FROM N.A.
X STRAIN=J1;
A Nicholas J.;
L Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
C -!- FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA
C REPLICATION.
C -!- SUBCELLULAR LOCATION: Nuclear (Probable).
C -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN
C FAMILY.
C
C This SWISS-PROT entry is copyright. It is produced through a collaboration
C between the Swiss Institute of Bioinformatics and the EMBL outstation
C the European Bioinformatics Institute. There are no restrictions on its
C use by non-profit institutions as long as its content is in no way
C modified and this statement is not removed. Usage by and for commercial
C entities requires a license agreement (see http://www.isb-sib.ch/
C or send an email to license@isb-sib.ch).
C
C R EMBL: U43400; AAC54703.1; -
C R PIR: T41943; T41943.
C R InterPro: IPR000635; Viral DNA bind.
C R Pfam: PF00747; Viral DNA bp; 1.
C W DNA-binding; DNA replication; Zinc-finger; Nuclear protein.
C T ZN_FING 459 475 C2HC-TYPE
C Q SEQUENCE 1131 AA; 129008 MW; 52697388D1B6D04F CRC64;
Query Match 7.7%; Score 483; DB 1; Length 1131;
Best Local Similarity 20.8%; Pred. No. 1.2e-28;
Matches 258; Conservative 204; Mismatches 560; Indels 220; Gaps 46;
2 ENTQKTVTPGPIGVYVACHVEVDLDEESFLAARSTDSDLLPLMRNLTVKFTSS 61
```

```
Db 3 DDNETVSAPICTAAWYILPKQKLEIITLTLSEMKRSWISPLLLNLTVENDSPFT 62
Qy 62 LAVVSGARTTGLAGAGITLKLTTSHFYPSVFPVHGGKGLVPSSA-APNLTRACNAARERF 120
Db 63 V-----KTFIIVGTVITKITSFNPVCFVFFHGTDFVFLKEADHGKLDKLCQTRKF 115
Qy 121 GFSRCQFPVGVAVETTG-ABICTRGLGEPENTILYLVTTALPKAVFMCMNVFLHYGLD 179
Db 116 NL-----QEFVNGNRKSDIGKICESVGRNADDVLCHIVVNGNGFKELLF-----AGLL 164
Qy 180 I-----VHNHGD--VIRIPLFPVQLFMPDVRNLVDPDPNTHRSIGEGFVPTFYNT 231
Db 165 IPCVEEQIQVQGECLAIKIPLYSATLFESEELCI-----DTCTEFIQENGFYAP 215
Qy 232 GLCHLIHDCVIAPMAVALVRNVNTAVARGAAHLAFDENHEGAVLPDITVTYFQSSSGT 291
Db 216 QISELVFLYIFTSGWMTLRFNNTLELIKAGLKQFIQDTEQTVKLAFNKTV-----265
Qy 292 TTARGARRNDVNSTKPSGSGFERRLASTMAADTALHAEVFN-----TGIYEETP-T 344
Db 266 --HGIPGQKLSPIEKD-----HMLVLD-AVITELTFSYTAAYLDSIYENQIM 310
Qy 345 DIKEMPFIQMEGTLPRLNALGSYTAHVAGVIGAMVFPSPNSALYLTVEDESGTEAKDGG 404
Db 311 NFSEWPIIKSAETHEEKIVELKKLRLHLSHVAALVFAANSILYSNKLAYISNTK-----365
Qy 405 PGPSFN-----RFYQAPAGPHLAANPQTRDGH-VLSQSTGSSNTEFVSVDYALIC 454
Db 366 --QAFNSAITQETLLRSIOFCNSLSINEDFYNDARKLIKNSSPCKEDKFAFLAYAC 423
Qy 455 GFGAPLLARLLFYLERCDAGATGGHGDALKVGTGTFDSEIPECSICEKHTRPVCAHTVH 514
Db 424 ATCPQILSHIINLNEWSIYNTNCGNSEIYNHIVNC--SSNLCEFCGCKCHSCITGTAII 481
Qy 515 RLQRMPRFGQATRQPIGVGTWNSQYSDCDPLGNAPYLYILRKPGDDTEAAKATMODTY 574
Db 482 RINSRLPQISKTTKKEPIVMTMFSRFYADVLDVLSFGKKGKVNESKDPNKEAQTPSLDRP 541
Qy 575 RATLERLFIDLEQERLLDRGAPCSSEGLSVIV--DHPTFR-----RLDTRLARIQT 626
Db 542 K-----FLGMTHDY-----CKQNLDALTGEDNLFKSNQDNFVNMINDLIQIEEA 588
Qy 627 TQQFMVLVETRDYKIREGLSEATHSMALTDPYSGAFCPITNFLVKRTHLAWVDLALS 686
Db 589 VSKCISEMERKTQ--TSREQIENCLOSFNIDTTPLSAFSPFPVYKYVILLVQLNAL- 645
Qy 687 QCHCVFYGOOVE-----GNFENQOPVLRRFVDLFNGGFISTRSITVTLSSEGPS 738
Db 646 ---IIGTVVDRPCTGNLISKWLMQOYOSLYGAFYNHSHFKKGLNMTVKI-----694
Qy 739 APNPTLQ-----DAPAGRTPDGLARVSVEVIRDIRVQRVVFSGNCTNLSEAR 789
Db 695 ASNVDMQEVIDFNLFKSGKYAKTSIQAKLCLSLWQCLADFAVKRPFNPKNTQNN---751
Qy 790 ARLVGLASAYQOEQRVDMHGLGFLIKQPHGLLFPFGMPNPSKSPNPQFWTLQRNQ 849
Db 752 -----PFFKVKQKQNPGLSGCLSFLFKYHERLF-----PNLKISCLE-FWQRIILNN 798
Qy 850 MP--ADKLUTHEBITT-IAAVKRFTEEYAAINFNLPPTCIGELAQFYMANLILKYCDHSQ 906
Db 799 MEKTDIGNVEDMRSPKFTFRVNSYDEIDLDIQECLLSFIETFYFNKLLSVLGYRD 858
Qy 907 YL--INTVTSITGARRPRDPSSVLHWRKDVLT--SAADIEIQAKALLEKTENLPELWTT 962
Db 859 YLTSLHALTSKLIV-----PQNPMLFPVFLKEHPTFSSVQEVVHVHVKLV--GNGLKEPMTA 912
Qy 963 APTSTHLVRAANNQRPVVLGISISKYHGAAGNN--RVFQAGNWSGLNGKNTVCPL-----1016
Db 913 SLTKEPNFGSFTGSIITFGIMTEKFVSVASRDYFHFQGLGWTAGSGVDNRNLNPPSSGL 972
Qy 1017 --FTFDRTRFIIACPRGGFICPVTPGSSGNRETTLSD--QVRGII-----VSGGAMVQ 1066
```

Db 973 QDFRMR-QKFIATK-...LCDII-VKKVKEATVDEVIRGKVLNIBESLSNYPE 1026

QY 1067 LAIYATVAVCARAQHMAFDMLSLTD-BELARDLE-...ELHDOIQTLETP 1116

Db 1027 LLILAENK-...DRDSKFTMDMLFYVDRPLAKSVNNKIQHTDNLNVHDFSLTLLSV 1083

QY 1117 W-...TVEGALEAVKILDEKTTAGDEPT 1142

Db 1084 FEEQVEDSAAYDFSELVEGNEQFGILKCEHEHEEFS 1125

RESULT 15

DNBI_SCMVC

ID DNBI_SCMVC STANDARD; PRT; 1160 AA.

AC P13215;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Major DNA-binding protein (MDEP).

GN UL27 OR DBP.

OS Simian cytomegalovirus (strain Colburn).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Betaherpesvirinae; Cytomegalovirus.

OX NCBI_TaxID=50292;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=91037979; PubMed=2172458;

RA Anders D.G.;

RT "Nucleotide sequence of a cytomegalovirus single-stranded DNA-binding protein gene: comparison with alpha- and gammaherpesvirus counterparts reveals conserved segments.";

RT J. Gen. Virol. 71:2451-2456 (1990).

RL [2]

RN SEQUENCE OF 205-308 FROM N.A.

RP MEDLINE=88155776; PubMed=2831398;

RA Anders D.G., Gibson W.;

RT "Location, transcript analysis, and partial nucleotide sequence of the cytomegalovirus gene encoding an early DNA-binding protein with similarities to ICP8 of herpes simplex virus type 1.";

RL J. Virol. 62:1364-1372 (1988).

CC -! FUNCTION: SINGLE-STRAND DNA-BINDING PROTEIN REQUIRED FOR DNA REPLICATION.

CC -! SUBCELLULAR LOCATION: Nuclear (Probable).

CC -! SIMILARITY: BELONGS TO THE HERPESVIRUSES DNA-BINDING PROTEIN FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC ENBL; D00750; BAA00647.1; -

CC ENBL; M19868; AAA46066.1; ALT SEQ.

CC InterPro; IPR000635; Viral DNA bind.

CC Pfam; PF00747; Viral DNA bp; 1

CC DNA-binding; DNA replication; Zinc-finger; Nuclear protein;

CC Early protein.

CC ZN FING 465 479 C2HC-TYPE.

CC SEQUENCE 1160 AA; 129005 MW; 7037716816974B1A CRC64;

CC

Query Match 7.6%; Score 479.5; DB 1; Length 1160;

Best Local Similarity 21.3%; Pred. No. 2,3e-28;

Matches 279; Conservative 202; Mismatches 57; Indels 255; Gaps 53;

Y 1 MENTQKTVTVTPGPGYVACVDELDLEISFLAARSTDSLALLPLMRNLNLTVEKFTS 60

Y 1 MGNEISALAPYCPAAYVYFTKTNHMEVIALSLCDSPPVVIAPLMGLTVDQDCT 60

Y 61 SIAVVSQARTTGLAGAGITLKTTHSHYFVSFVHGGKHVLP-SAAAPNLTRACNAARER 119

Db 61 SV-...RTPVVCVGGVGLTKVTS-PCFFALYFVNTQGVDFSEPHGDVQRLCDETRQR 113

QY 120 FGFSRCQPPVDGAVETTGAEICTRLGLEPENTILYLVVLTALFKEAVFMCNVLHYGLD 179

Db 114 YATESVM-PEEGRAPTDLAALCTAAGCDPQBLVHVVGNGMEFMYAGOLIFCFEEAA 171

QY 180 IVHINHGIVIRIPLFPVOLP--MPDVNRLVDPFNTHRSIGEGFVYPTFYNTGLCHL 236

Db 172 PTRLNDCDAVRVLPYPTLFGSLQADV-...SDELSLDKES-...SFVESRGLYVPAYSET 225

QY 237 IHDCVIAPMVALVRNVTVAVARGAAHLAFDENHEGAVLPDITITYFQSSSSSTTTARG 296

Db 226 LFYYVYVTSWCQALRFSETKVLEAALKQFVNDSSQSVKLAPHKKYFGYTSOK-... 277

QY 297 ARRDVNSTKPSGSGPERRLASIMAADTAL-...HAEVFNPTGIYBETPTDIKWP 350

Db 278 -...LSSLEK-...HMLSDAIVICLGFSGFASVFLDSAYGASDSMVYSEWP 321

QY 351 MFIGMEGTLPRINALGVSYARVAGVIGAMVFSNSAL-...YLTEVEDSGMTEAKDGG 404

Db 322 VVVATDHRDLIRALTELKHLSTHISALLFSCNLSILYHNRLLVLTNSKNASGT-...GA 377

QY 405 PGPSFNRYQFAGPHLAANPQTDGSH-VLSSQSTGSSNTEFSVDYIALICGFCAPLLAR 463

Db 378 SQEVLKSIHFANGLTGLCEDTYNDARKLKCGVWAKDERYAPYHLSLICGTCPCPLFSA 437

QY 464 LLFYLERCDAFTGCHGDALKYVTGTFDSE-...IPC-...SICEK-...HTRPVCA 509

Db 438 FIWLNRVSV-...YNTGLTGSSTLSNHLICGSSLCGACGTCCHT-...CY 482

QY 510 HTTVHRLRQMRPFGQATROPIGVFGTMNSQVDCDPLGNAPYAPYLILRKPQDQTEA-... 565

Db 483 NTAFRVQTRLPQMPRLPKKPSVVMQSRFLNDVDVLGTFG-...RRYSAESKEASLDA 538

QY 566 -...AKATMQTYRATLERLEIDLEQERLLDRGAPCSSEGLSSVVD-HPT 611

Db 539 KADBSASTSNRTASSSVDRTHR-LNRILDYCKQWELID-...SVTGEDTWTNGRSD 591

QY 612 FRRILDTLRARIEQTITTFQMKVLVETRDYKIREGLSEATHSMALTDPYSGAFCPITNEL 671

Db 592 FINLVSSLNKFDDEAMSFVS-...EVRMKSNDREVLGATQAFNLDLNPFAVSFPLAYE 648

QY 672 VKRTHLAVODLALSCHQCVFYQC-...QVEGNFRNQFPVLRFRFVDFLNGGFTSRS 727

Db 649 YYRVIFAIIONVALITATSYIVDNPLTTSLSVRWVWTFQPSI-...HGAFSTSS 699

QY 728 ITVTLSEGPVSAENPLTGODAPAGR-...TFDGLARVSVVEVIRDIRVK 772

Db 700 -...RKGFIFIRNVKSKNADHRLDPLFKLYARGTYSVISMEIKLSRUSVPSLLNFRVK 754

QY 773 NRVVGNGCTNLSEAAARLVGLASAYORQE--KRVDMHLGALGFLLKQFHGLLPRGMP 830

Db 755 NRPI-...SKASKGT-...TAHVFFRRRHVPKCNPKVKGCLGFLLYKXVKLFP-... 799

QY 831 PMSKSNPQWFMTLLQONQMPADKLTHEITIAVKRF-...TEEYAAININLPP 883

Db 800 -...DCGFSCLQFQKVCANALP-...KNVNIQMGPFNNFVKFVISVTADYNEDHIDVPP 853

QY 884 TCIQELAQFYMANILKYCDHSQYL-...INTLTSIITGARRPRDPSSVLHWIRKWDVTSAD 941

Db 854 DCMNLNLENRHNKFLCFYGFQDYIGTLHLGLTTLTYQNHAQFP--YLLGESPNFASAD 911

QY 942 IETQAKALLEKENIPELMTTFTTHLVRAAMNQRPMVVLGIS-SKYHGAAGNNRVFQA 1001

Db 912 FALRLKDL-KATGVTAPLASTVTRESLMRTIFEQRLSVTSFSIEKYAGVNNKKEIYQF 969

QY 1002 GN-...WSG-...GGKNVCPLFTDRTERRFIACPRGGFCIPVTGPGSGNET 1048

Db 970 GQIGYFSGNGVRSINTNSIGQD-...YKFWR-ORCLATKLSVLI-...KSRSDN 1019

QY 1049 TISDQVRGIIVSGGAMVOL-...AIVATVAVCARAQHMAFDMLSLTD- 1095

Db 1020 VLFDE-...DIKRWAAALSDENLDVDPMLMAY-...EILSTREEIPERDDVLFVDGC 1072

```

1096 EFLARDLEELHDQIIQTLTPWTVGALAVKILDEKTTAGDGETPTNLAEPDCEPSSH 1155
1073 QAVADSLMEKFSRLOEMGVDDFSL---VNLOQLDSRPECGGGGEV-----H 1117
1156 DTTSNVLNISGNSIGSTVPGK-RPPED--DELFDLSGIPKHN 1199
1118 DLSALFTAASGEAVGNSV--GLNARGGSHAFDE--DCOLLFAKGR 1160

```

search completed: January 30, 2004, 13:13:04
 OS time : 22 secs

GenCore version 5.1.6
Copyright (C) 1993 - 2004 CompuGen Ltd.

MM protein - protein search, using sw model

run on: January 30, 2004, 13:10:47 ; Search time 27 Seconds
(without alignments)
4284.849 Million cell updates/sec

Title: US-09-769-699-2

Effect score: 6294
Sequence: 1 MENTQKTVTGTPLGYVYA.....DEFLDGLPIKHNITWEM 1203

Coring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR 76.*

- 1: PIR1.*
- 2: PIR2.*
- 3: PIR3.*
- 4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6294	100.0	1204	1 DNB29	DNA-binding protei
2	3620.5	57.5	1209	1 DNBEC4	DNA-binding protei
3	3583	56.9	1208	2 T42574	DNA-binding protei
4	3054	48.5	1196	1 DNBKS	DNA-binding protei
5	3049	48.4	1196	1 DNBV1	major DNA-binding
6	3047	48.4	1196	1 DNBHF	DNA-binding protei
7	3003	47.7	1197	1 A48350	DNA-binding protei
8	2933	46.6	1186	1 DNBEG	DNA-binding protei
9	1094.5	17.4	375	2 JQ0846	DNA-binding protei
10	669	10.6	1128	1 Q0B847	DNA-binding protei
11	633	10.1	1128	2 T42922	major single-stran
12	618	9.8	1145	2 S55600	single-stranded DN
13	604	9.6	1128	1 DNBEM1	DNA-binding protei
14	596	9.5	1127	2 T03105	major single-stran
15	490.5	7.8	1132	2 T44001	major single-stran
16	483	7.7	1131	2 T41943	major DNA-binding
17	479.5	7.6	1160	2 A36256	DNA-binding protei
18	477	7.6	1235	1 Q0BEM4	DNA-binding protei
19	412	6.5	1191	1 A44051	DNA-binding protei
20	237.5	3.8	483	2 S69894	major DNA-binding
21	137	2.2	1504	2 T17426	FK506 polyketide s
22	120.5	1.9	781	2 A13014	methyl-accepting s
23	120.5	1.9	788	2 F98269	hypothetical prote
24	119.5	1.9	701	2 A11501	internalin protein
25	117.5	1.9	358	2 H83380	O6-methylguanine-D
26	116.5	1.9	335	2 H81702	adherence factor T
27	115.5	1.8	756	2 T17428	FK506 polyketide s
28	115	1.8	823	2 B35963	protein-tyrosine k
29	115	1.8	2505	1 XYRTFA	enoyl-[acyl-carrie

ALIGNMENTS

RESULT 1

DNBE29

DNA-binding protein - human herpesvirus 3

C:Species: human herpesvirus 3, varicella-zoster virus

C:Date: 30-Sep-1988 #sequence_revision 30-Sep-1988 #text_change 16-Jul-1999

C:Accession: C27214

R:Davidson, A, J, Scott, J, E

J:Gen: Virol. 67, 1759-1816, 1986

A:Title: The complete DNA sequence of varicella-zoster virus.

A:Reference number: A27345; MUID:86306657; PMID:3018124

A:Accession: C27214

A:Molecule type: DNA

A:Residues: 1-1204 <DAV>

C:Cross-references: EMEL:X04370; NID:G59989; PIDN:CAA27912.1; PID:G60018

C:Genetics: 29

C:Gene: 29

C:Superfamily: herpesvirus DNA-binding protein

C:Keywords: DNA binding

Query Match	100.0%	Score 6294	DB 1	Length 1204
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1203	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MENTQKTVTGTPLGYVYACRVEDLLEISFLAARSTDSDLALLPLMNLTVKFTFS	60	
DB	1	MENTQKTVTGTPLGYVYACRVEDLLEISFLAARSTDSDLALLPLMNLTVKFTFS	60	
QY	61	SLAVVSGARTTGLAGAGITLKLTTSHFYPSVVFHGGKHLVPSAAPNLTRACNAARERF	120	
DB	61	SLAVVSGARTTGLAGAGITLKLTTSHFYPSVVFHGGKHLVPSAAPNLTRACNAARERF	120	
QY	121	GFSCQGGPPVDGAVETTGACICTRLGLEPENTILYLVTALFKAEVFMVNFVLYHGGLDI	180	
DB	121	GFSCQGGPPVDGAVETTGACICTRLGLEPENTILYLVTALFKAEVFMVNFVLYHGGLDI	180	
QY	181	VHINHGDIIRIPFPVOLFMFDPVNRVLPDPENTHRSIGSGFVYPTFYNTGLCHLHDC	240	
DB	181	VHINHGDIIRIPFPVOLFMFDPVNRVLPDPENTHRSIGSGFVYPTFYNTGLCHLHDC	240	
QY	241	VIAPMAVALRVNVTAVARGAHLAFDENHEGAVLPDITTYTFQSSSGTTTARGARN	300	
DB	241	VIAPMAVALRVNVTAVARGAHLAFDENHEGAVLPDITTYTFQSSSGTTTARGARN	300	
QY	301	DVNSTKSPSGGFERRLASIMADTALHAELVFNTEPTDIKEPMFTIGEGTLP	360	
DB	301	DVNSTKSPSGGFERRLASIMADTALHAELVFNTEPTDIKEPMFTIGEGTLP	360	
QY	361	RLNALGSYTVARVAGVICAMVFSNSALYTEVSDSGMTAKDGGPGFSNRFYQFAGPHL	420	
DB	361	RLNALGSYTVARVAGVICAMVFSNSALYTEVSDSGMTAKDGGPGFSNRFYQFAGPHL	420	
QY	421	AANPQTRDGHVLSQSSNTSEFSDVILALICGGFAPLALRLLFYLERCDAGATGGH	480	

09/769,699
2-21-04
Search Notes

```

Db 421 AANPQTDGHLVLSQSSNGSTSSNTEFSDVYLALICGFGAPLARLLFYLERCDAGFTGGH 480
QY 481 GDALKYVGTGDSSEIPCSLCEKHTRPVCAHTTVHRLRQMRPFQOATRPQIGVGTNSQ 540
Db 481 GDALKYVGTGDSSEIPCSLCEKHTRPVCAHTTVHRLRQMRPFQOATRPQIGVGTNSQ 540
QY 541 YSDCDPLGNVAPYLTILRKPGDTAAKATMDQTYRATLERLFDLEOERLLDRGAPCSSE 600
Db 541 YSDCDPLGNVAPYLTILRKPGDTAAKATMDQTYRATLERLFDLEOERLLDRGAPCSSE 600
QY 601 GLSSVVDHPTRFRILDLTRARISQTTQTFQMKVLVETRDYKIREGLSEATHSMALTFDPY 660
Db 601 GLSSVVDHPTRFRILDLTRARISQTTQTFQMKVLVETRDYKIREGLSEATHSMALTFDPY 660
QY 661 SGAFCPITNFIKVLTHLAVODLALSCHQCVFYGOQVEGRNFRNQFVLRPFVDLPNG 720
Db 661 SGAFCPITNFIKVLTHLAVODLALSCHQCVFYGOQVEGRNFRNQFVLRPFVDLPNG 720
QY 721 GFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTPDGLARVSVEVIRDKVRNVFSGN 780
Db 721 GFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTPDGLARVSVEVIRDKVRNVFSGN 780
QY 781 CTNLSEARARLVGLASAYQOEKRVDMHLGALFLKQPHGLLFRGMPNPKSPNPQW 840
Db 781 CTNLSEARARLVGLASAYQOEKRVDMHLGALFLKQPHGLLFRGMPNPKSPNPQW 840
QY 841 FWTLLQRONMPADKLTHEETTTAAVKRFEETEEYAAINFNLPTCIGELAQFYMANLILK 900
Db 841 FWTLLQRONMPADKLTHEETTTAAVKRFEETEEYAAINFNLPTCIGELAQFYMANLILK 900
QY 901 YCHSQYLINTLSITGARRPDPSVLHWIKDVTSAAIDTQAKALLEKTENLPELW 960
Db 901 YCHSQYLINTLSITGARRPDPSVLHWIKDVTSAAIDTQAKALLEKTENLPELW 960
QY 961 TTAFTSTHLVRAAMNORPMVLGISIKVHGAGNNRVFOAGNWSGLNGKNCVCLFTPD 1020
Db 961 TTAFTSTHLVRAAMNORPMVLGISIKVHGAGNNRVFOAGNWSGLNGKNCVCLFTPD 1020
QY 1021 RTRRFIACPRGFCFVTPGSSGNRETTLSQDVRGIIYSGGAMVOLAIYATVRAVGAR 1080
Db 1021 RTRRFIACPRGFCFVTPGSSGNRETTLSQDVRGIIYSGGAMVOLAIYATVRAVGAR 1080
QY 1081 AQHMAFDDMLSLTDDLEFLARLDELHQIIOLETPTWTEGALBAVKILDEKTTAGDGET 1140
Db 1081 AQHMAFDDMLSLTDDLEFLARLDELHQIIOLETPTWTEGALBAVKILDEKTTAGDGET 1140
QY 1141 PTNLAFNFDSCPSHDTTNSVLNLSGNSISGTSVPGKRPDDDELFDLSGIPIKHGNIT 1200
Db 1141 PTNLAFNFDSCPSHDTTNSVLNLSGNSISGTSVPGKRPDDDELFDLSGIPIKHGNIT 1200
QY 1201 MEM 1203
Db 1201 MEM 1203

```

RESULT 2

NBECA

NA-binding protein - equine herpesvirus 1 (strain Ab4p)

Species: equine herpesvirus 1

Note: host Equus caballus (domestic horse)

Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999

Accession: E36798

Submitted to GenBank, March 1992

Tel: 1-800-541-5300, M.S.; McBride, K.; Davison, A.J.

Description: The DNA sequence of equine herpesvirus-1.

Reference number: A36805

Accession: E36798

Molecule type: DNA

Residues: 1-1209 <TEL>

Cross-references: GB:M86664; NID:G330791; PIDN:AAB02466.1; PID:G330823

Tel: 1-800-541-5300, M.S.; McBride, K.; Davison, A.J.

irology 189, 304-316, 1992

A;Title: The DNA sequence of equine herpesvirus-1.
 A;Reference number: A41831; MUID:9229566; PMID:1318606
 A;Contents: annotation; possible protein-coding frames
 A;Note: neither amino acid nor nucleotide sequence is given
 C;Genetics:
 A;Gene: 31
 C;Superfamily: herpesvirus DNA-binding protein
 C;Keywords: DNA binding

```

Query Match 57.5%; Score 3620.5; DB 1; Length 1209;
Best Local Similarity 57.1%; Pred No. 1e-266;
Matches 698; Conservative 191; Mismatches 299; Indels 35; Gaps 14;

QY 1 MENTOKTVVTGPGYVYACRVEDDLLEEFISFLAARSTDSDLALLPLMRNLIVKFTFTS 60
Db 1 MESAPKTVSLPVSPLGYVYARQKASLQGTGTVSLTAARSVDSDLAVLVIRGLTVEQFTT 60
QY 61 SLAVVSGARTTGLAGAGITLKLTTSHFVPSVFPVPHGKHVLPSSAANLTRACNAARERF 120
Db 61 NVAVVAGSKTTLGLGTTLKLTPSHFNPAFYGGSVIGASSNAENLTRACEAARRF 120
QY 121 GFSRCQPPVDGAVETTGAETCTELGLEPENTIIYLVVLTALFKEAVFMCNVFLHYGLDI 180
Db 121 GFSAFSSPPVENAVETSGEBCASLNSPETTALYLVVTSFKEMVYVCTVFLHYGSTST 180
QY 181 VHINHGDIPIPLFPVOLFMVDVNRVLPDPFNTHRSIGEGFVYPTPTNYTGLCHLHDC 240
Db 181 VTIDQDQAMKPIYPVOLYMPDVNRVLPDPFNTHRSIGEGFVYPTPTNYTGLCHLHDC 240
QY 241 VIAPMAVALRVNVTAVARGAHLAFDENHEGAVLPDITVTVTFOSS--SSGTTTARGAR 298
Db 241 VLGPAAVALRVNLDVGARGAHLADENHEGAVLPDITVTVTFOSS--SSGTTTARGAR 298
QY 299 RNDVNTSKSPSGGFRERRLASINAAADTALHAENVFNTGIYETPTDIKEWPMFIMGEGT 358
Db 301 QGD--GSGSKNSASGIERRLASVAAADTALSVDISIMGAGIYDTLPSEVDPVLSGDDT 359
QY 359 LPRNALGSTRVAVGVIGAMVSPNSALVLTVEVDSGMTAKDGGGFSFNRFYFAGP 418
Db 360 -ESLEALGCAVRLSLGLVGMVANSVLTVEVDSGMTAKDGGGFSFNRFYFAGP 417
QY 419 HLAANPOTDRGHVL-----SSQSTGSSNTEFSDVYLALICGFGAPLARLLFYLERCD 472
Db 418 YVAGNPOTDRGHVLPHATDQQAAPINGNSQBFSLDYALACGFCQILARLLFYLERCD 477
QY 473 AGAFTG-GHGDAKYVTGTFDSEIPCSLCEKHTRPVCAHTTVHRLRQMRPFQOATRPQI 531
Db 478 AGTGGNETDAURYLANLTLESVPCGLCNOATRPACATTLHRLRQMRPFQOATRPQI 537
QY 532 GVFGTMNSQYSDCDPLGNVAPYLTILRKPGDTAAKATMDQTYRATLERLFDLSQERLL 591
Db 538 GIFGTMNSAYSDCDPLGNVASYGALKRPND-NEAPKSIMQDTYRATMERLVELSQAKLI 596
QY 592 DR-----GAPCSSEGLSSVIVDHPTRFRILDLTRARISQTTQTFQMKVLVETRDYKIREGL 646
Db 597 DKETLAQASPCAP--TSVVDHQAQSFGLLSNIKDTIEGAARQFMTLVEARDFKIREGL 654
QY 647 SEATHSMALTFDPYSGAPCPITNFIKVLTHLAVODLALSCHQCVFYGOQVEGRNFRNQF 706
Db 655 ADANHTMSILDFPYSSTFCPTVSLARRTVFVLQDLVLSQCHCLFVQSGVEGRNFRNQF 714
QY 707 QPVLRFRFVLDLPNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTPDGLARVSVEVI 766
Db 715 QPVLRFRFVLDLPNGGFISTRSITVTLSEGPVSAPNPTLGQDAPAGRTPDGLARVSVEVI 774
QY 767 RDIKVRNVVFGSNCNTNLSAARARLVGLASAYQOEKRVDMHLGALFLKQPHGLLFP 826
Db 775 RDIKVRNVVFGSNCNTNLSAARARLVGLASAYQOEKRVDMHLGALFLKQPHGLLFP 834
QY 827 RGMPPNPKSPNPQWFWTLLQRONMPADKLTHEETTTAAVKRFEETEEYAAINFNLPTCIG 886
Db 835 RGMPPNPKSPNPQWFWTLLQRONMPADKLTHEETTTAAVKRFEETEEYAAINFNLPTCIG 894

```



```

QY 897 GELAQFYMANLILKYCDHSQYLINTLIIITGARRPRDPSSVLHWIRKDVTSAAADIETQA 948
Db 895 GELAQFYFANLVLKYCDHSQYFINGLTAIVUGSRPRDPAVLAWTDRTINGAADVEPAA 954
QY 947 KALLEKTENLSELTWTAFTSTHLVRAAMQORPMVVLGISISKYHGAAGNNRVFOAGWSG 1006
Db 955 QEVLRQSGNFAAATGFTTNNMRYVMYDMQRPVWVIGLSISKYNSAGNNRVFOAGWNG 1014
ZY 1007 LGGGNQVCPLEFTRRRFIIACPRGGFICPVTGPGSGNNRETTILSDVRGIITVSGAMVQ 1066
Db 1015 LGGGNQVCPLEMAFDTRRRFVLACPRVGTCEAGFGTGVRENTLSEQVRGIVSEGGPMVQ 1074
ZY 1067 LAIVATVVRVARGAQHAWDDMLSLTDBFLARDLEEJHDOIITLTPMTVEGALEAV 1126
Db 1075 TAVFAAVLHLAGARTQHLAVDDMIGLWDDDEFLAASLDALNATWVDQF-GEWSVEAAQELV 1133
ZY 1127 KILPEKTTAG---DGETPTNLAFNFSQ---EPSHDTTSNVLNISGNSISGTVPELKR 1180
Db 1134 KMEAAQTAGVAAGG---ADFQACVGDTPQOSTSA----FNGGLMAAAPAGQKRS 1185
ZY 1181 PEDDELEFDLSGIPKIGNITNEM 1203
Db 1186 LPDDILFDMGAPPEKKSGLTFDM 1208

RESULT 3
42574
NA-binding protein - equine herpesvirus 4 (strain NS80567)
;Species: equine herpesvirus 4
;Variety: strain NS80567
;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
;Accession: T42574
;Telford, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.
;Title: The DNA sequence of equine herpesvirus-4.
;Reference number: Z22173; MUID:98264497; PMID:9603335
;Accession: T42574
;Status: preliminary; translated from GB/EMBL/DBJ
;Molecule type: DNA
;Residues: 1-1208 <TEL>
;Cross-references: EMBL:AF030027; NID:g2605950; PIDN:AACS9547.1; PID:g2605975
;Experimental source: strain NS80567
;Genetics:
;Gene:31
;Superfamily: herpesvirus DNA-binding protein
;Keywords: DNA binding

Query Match 56.9%; Score 3583; DB 2; Length 1208;
Best Local Similarity 56.4%; Pred. No. 7.4e-264;
Matches 689; Conservative 194; Mismatches 305; Indels 34; Gaps 13;

/ 1 MENTOKTVTVPGTPIGYVYACRVEDLDLEEISFLAARSTDSDLALPLMRNLTVKTFTS 60
> |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
/ 1 MESAPKTVSLPSPPLGYVVAIQNTFMETEAALTMAARSIDSDLAVLPVIRGLTVEQTFTT 60
> |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
/ 61 SLAVVSGARTTGLACAGITLKTTSHPVSPVVFHGHKVLPSAAPNLTACNAARERF 120
> |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
/ 61 NVAVVAGKTTGLGAGATTLLTTSHTPEAFVFGGVSFCASSKAPNLTACELARRRF 120
> |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
/ 121 GFSRCQGPVDPGAVETTGAECTRLGLEPNTILYLVTALEAVEMCNVFLHVGGLDI 180
> |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
/ 121 GFSPPSPFPDNAVETSGEEICASINLSPETTTLLVVTTFKMYMNCNTFLHVGGTST 180
> |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
/ 181 VHINHGDIVRPLPEVOLFMEDVNRVLDPDNTHRSIGBGFVPTFPYVNLGLCHLHDC 240
> |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
/ 181 VTIHQEAQVAKIPIPVQVLPDVRNLAEPFNKSRKSIDGEPFYYSKFPFNSDLCLLHGY 240
> |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
/ 241 VIAPMAVALRVNVTAVARGAAHLAFDENHEGAVLPDDITYTVFQSS--SSGTTTARGAR 298
> |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
/ 241 VLGPAVALRVNLGDVARGAAHLADENHEGSVLQDVTFLFDSAQTSKSGSGRTQR 300
> |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
/ 299 ENDVNSTSKPSPGGERLASIMADTAUHAENVNTGIYSETPTDIKEWPMFIGEGT 358
> |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::

```

Db	301	QGD- GSLKXNGSSGIERRLASIMAA DTALSVDISMGAGVDTLP SVBDLPI-LSVGDD	358
Qy	359	LPRLNALGSYARVAGVICAMV FSPNSALYLTEVDSGMTEA KDGPGPSFNRFFYQFAGP	418
Db	359	RERLEALGAYARLSGLVGMV FSANSVLYMTVEVDGPGAD KGDAS-NPSYHRFLLIAP	417
Qy	419	HLAANPOTDRDGHVLS----- SQSTGSGNTSFSDVLYALIC GFGAPILARLLFYLERCD	472
Db	418	YVAGNPOTDKGRVLIQHTADQ APAAPINGSNQBSLDYALAC GFCPLLARTLFYLERCD	477
Qy	473	AGAFTH-GHGDAKLVYGTGTF DSEIIFCSICEKHTRPVCAT HTVHRLQRMRFQATROPI	531
Db	478	ACTFGGRNETDALRYLANL TESEVFCGLCTPATRPACAT HTLHRLQRLPRFGTPVRAP	537
Qy	532	GVFGTWNISYSDCDPLGNVAP YLIILKFGDQTEAAKATWQD TYRATILBRLLFIDLEQERLL	591
Db	538	GIPTWNISYSDCDVLGNVASY GALKREND-NEAPKSIQD TYRATIMERLVNDLEQAKLI	596
Qy	592	DRGA-----PCSSGLSVIV DHPHTFRILDTLAARIEQTT TFQMKVLVEYRDVKYIREGL	646
Db	597	DREALAHAGTCSAS--TG VWKDQASFINLLSTIKDIT EGAAEQFMTLVEYRDKPIREGL	654
Qy	647	SEATHSMALTPDPYSGARCP TINELVRETHLAVQDLALS OCHCVFQGVQVQVEGRNFRNQF	706
Db	655	ADAWNTMSISLDPYSSFC PVTSELSRTIFAVLQDLVL SOCHCLFTGQSVVEGRNFRNQF	714
Qy	707	QVLRERRFVDFLNGGFI STRISITVTLSEGPVSA PNPTLGQDAPAGRTFDGDLARVSVEVI	766
Db	715	QPVLRRLPLDMLNGGFI TAKTVTVVSDSGVTAENL TLPSSEPTKDYDGDMAVRSNEVL	774
Qy	767	RDIRVKNRVFSGNCTNL SEARARLVGLASAYORQ EKRVDMHLHGALGFTLLKQPHGLLFP	826
Db	775	RDLRKRVLFSNGANSE ARARVAGMASVRRPEK GNIINGAVGFLVKQPHKVLFP	834
Qy	827	RGMPNKSQPNQFWTLL QKNQMPADKLTHEETIT TAAVKRFTVEYAAINFLNPTCI	886
Db	835	RGHPFGIDTPNQFWTLL QKNQMPARLLSKEDLET ITAIKRSHESYAINFLNFPNI	894
Qy	887	GELAQFYMANILKYCH DSOVLNLTLSITGARR RDPSSVLHWIKOVTSAA DIETQA	946
Db	895	GELAQFYFANILKYCH DSQVFNGLTAVVGGRR RDPAAVLAWNRTINGAS DVEPAA	954
Qy	947	KALLEKTENPELWTTA FTSHLVRAAMNQPMVL GISISKYHGAAGNNRVFQAGNWSG	1006
Db	955	QEVLLQQLGSGNPA AATGTFASTNMVRYMD ORPMVVLGLSISKYSGA NNRVFQAGNWSG	1014
Qy	1007	LNGGKNCVPLTFDRTR RFIIACPRGFCPTVGP SSGNRETLSQVIRGIIVSGAMVQ	1066
Db	1015	LNGGKNCVPLWADRTR RFVLCPRVGTCEAGFG MGARENLTSEQISIVSDG GPMVQ	1074
Qy	1067	LAIYATVVRAGARAC MAPDWSLSTDDFTL ARDEELHDOITQLET PWTVEGALEAV	1126
Db	1075	TAVFSVVLITAGART OHLAVDDWIGLVDD EFLAASLDALNAVWDQF- GEWSVEAAQDVI	1133
Qy	1127	KILDEKT-----TAG DGETPTNLAFNDSCE PSHDTSNVLNISG SNITSGSVPLGRPP	1181
Db	1134	RTWDAQTNNGVSTG G-----APDFGACV GDANQSSITTFNMGFA-- SSSAPAGQKRPH	1185
Qy	1182	EDELFDLSGIPKIGNI TNEM	1203
Db	1186	PDDILPDMGAPPEK SGLTFDM	1207

RESULT 4

DNBKS

DNA-binding protein - human herpesvirus 1 (strain KOS1.1)

C:Species: human herpesvirus 1

A:Note: host Homo sapiens (man)

C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999

C:Accession: A28601

RESULT 4

RESULTS AND DISCUSSION

DNA-binding protein - human herpesvirus 1 (strain KOS1.1)

C;Species: human herpesvirus 1

A; Note: host Homo sapiens (man)

C;Date: 31-Mar-1990 #sequence r

C;Accession: A28601

R: Gao, M.: Bouchev, J.: Curtin

Kisao, M.; Boucley, J.; Curtin, Virology 163: 319-329, 1988

VIROLOGY 163, 319-329, 1988

A;Title: Genetic Identification

[illegible]

176 GGLDVIHNGHVDIRIPFPVQLFMPDVRNRPDPENTHRSIGRGVYPTFPYNTGLCH 235
 179 GGSCKVTIGAEVHRIPVYQLFMPDPSRVABFNHRISGKFTYFPFFNRPNNR 238
 236 LTHDCVIAPMAVALRVNVTAVAGAAHLAFDENHGAFLVLPDITTYTFQSSSSSTTTAR 295
 239 LLFEAVVGAVALRCRNVDAVARAAHLAFDENHGAALPADITFTAFASQG--KTPR 296
 296 GARRNDVNSTKPSGSGGPERLASIMAAADTALHAELVFNIGIYETDIDKEWPMFQM 355
 297 GGR-----DGGKGAGGFEQRLASVWAGDAALALESVMAVDFEPTDISAMPLFEGQ 351
 356 EGTLPRLNALGYSYARVAGVIGAMVFPNSALYLTEVEDSGMTAKDGGPSPFNRPYQF 415
 352 DTAARANAVAGAYLARAAGLVGMVFNSTGALHLETVDDAGPADPKHSHK-PSFYRFLV 410
 416 AGPHLAANPQDRDGHVL-----SSOSTGSSNTEFSDVILALICGFCAPLLARLLFYL 468
 411 PGTHVAANPQDRGHVVPFEGPPTAPLVGGTQ--EPAGEHLAWLCCGFPALLAKMLFYL 469
 469 ERCDAGFTGGHG--DALKYVTGTDFSEIPCSICEKHTRFVCATHTVRLRORMPRFGOAT 527
 470 ERCDGAVIVGQEMDVFRYVADSNQDTPCNLCFTDIEHACVHTTLMRLRAHHPKFSAA 529
 528 ROPIGVGTWMSOYSDCDPLGNAPYLILRKPQGTAEAKATMODTYRATLERLFDLEQ 587
 530 RGALGVGTWMSMYSDCDVLGNVAAFSALKR--ADGSETARTIMOETRYAATERVMALET 588
 588 BRLLDRGAPCSSEGLSSVIVDHPTRFRLDLRLARIEQTTTQPMKVLVETRDYKIREGLS 647
 589 LQYVDQAVPTANGRLTIIITREALHTVVNNVQVVDREVEQLMRNLVEGRKFKFDGLG 648
 648 EATHSMALTDPYSGACFPITNFIKPKPDGTAEAKATMODTYRATLERLFDLEQ 707
 649 EANHMSLTLDPYACGCPCLQLLGRSNLAVQDLALSCQCHGVFAGQSVGEGNFRNQF 708
 708 PVLRRRVDFLNGGFISTRITVTTLSEG--PVSAFNPTLGDAPAGRTFDGLARVSEVI 766
 709 PVLRRVDMFNNGFLBAKTLVLSGAAICAPSLTAGOTAPAESFEGDVARVTLGFP 768
 767 RDIVKRVVFSNGCTNLSEARARLVGLASAYORQEKRVDMHLHGALGFLKQPHGLRPP 826
 769 KELRVKRVLFAGASANASAAKARVASLOSAYQPKRVVDILLGFLGLKQPHAAIPP 828
 827 RGMPNKSFPNQFWTLQGNQPADKLTHEITTTAAVKRFTTEYAAINFINLPPTCI 886
 829 NGKPPGNSQNPQFWTALQNLPARLLSREDIETIAFKFSLDYGAINFNLAPNV 888
 887 GELAQFYMANLILKYCDHSOVLINTLTSIITGARRRDPSSVLHWIRKDVTSADLETQA 946
 889 SELAWYTMANQLLYCDHSYFINLITAIAGRRRPPSVQAAA--SAQGGAGLEAGA 945
 947 KALLEKTENIPELMTTFTSTHLVRAAMNORPMVVLGISISKYHGAAGNVRVQAGNWSG 1006
 946 RALMDAVDAHFAGAWTSMFASCNLLRPVMAARPMVVLGLSISKYVAGNDRVQAGNWS 1005
 1007 LMGKNVCPLFTDRTERFIIACPGFICPVTPGSGNRETTLSQVRGIIVSGGAMVQ 1066
 1006 LMGKNVACPLIIDRTKFKVIACTPRAGFVCAASSLOGGAEHSSCLBQLRGIIISGGA 1065
 1067 LAIYATVVRVAGARAHMADDDWLSITDDBFLARDLEELHDQIIQLETPTWTEGAL--- 1123
 1066 SSVFVATVKSILGPRTOQLQIEDNLALLEDEYLSSEMMLTARALENGENWSTDALEVA 1125
 1124 -EAVKIIDEKTAGDGETPNNLAFND--SCFSPHDTTGNVLNIGSNISGTVGPKLKP 1180
 1126 HEALVSLQGNAGE-----VFNFGDOCE-----DDNATPFGGPGAPGAPAGRKEA 1173
 1181 PEDDELFDLSGIIPIKGNITMEM 1203
 1174 FHGDDPG-EGPDPKKGDLTDM 1195

RESULT 6

DNBEHF
 DNA-binding protein - human herpesvirus 1 (strain F)

C;Species: human herpesvirus 1

C;Note: host Homo sapiens (man)

C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 07-Jun-1996

C;Accession: D29242

R;Hammerschmidt, W.; Conraths, F.; Mankertz, J.; Pauli, G.; Ludwig, H.; Buhk, H.J.

Virolgy 165, 388-405, 1988

A;Title: Conservation of a gene cluster including glycoprotein B in bovine herpesvirus 1

A;Reference number: A94381, MUID:88306231, PMID:2841793

A;Accession: D29242

A;Molecule type: DNA

A;Residues: 1-1196 <HAM>

A;Cross-references: GB:M21629

C;Superfamily: herpesvirus DNA-binding protein

C;Keywords: DNA binding

Query Match 48.4%; Score 3047; DB 1; Length 1196;
 Best Local Similarity 49.8%; Pred. No. 4.7e-223;
 Matches 609; Conservative 199; Mismatches 367; Indels 48; Gaps 17;

QY 1 MENTQKTVT---VPTGPGLYVY--ACRVEDLDLEISFLAARSTDSDLALLPLMRNLVE 55
 Db 1 METKPKTATTIKVPPGLGYVYARACPSGIEL--LALLSARSGSDVAVAPLVGLTVE 58
 QY 56 KTFSTSLAVVSGARTGLAGGITLKTTSFYPSVVFHGGKHVLPSSAANPLTRACNA 115
 Db 59 SGFEANVAVVGSRTVGLGTAVSLKLPESHYSVSVYFHGGRHLDPSQTAPNLTRLCER 118
 QY 116 ABREFGFCRCQPPVDGAVETTGAEICTRLGLEPENTLYLVVLTALFKEAVFMCNVELHY 175
 Db 119 ARRHGFGSDYTPRGDLKXHTTGALCERGLDDRRALLVLTGEGKEAVSINNTFLHL 178
 QY 176 GGLDIVHNGHVDIRIPFPVQLFMPDVRNRPDPENTHRSIGRGVYPTFPYNTGLCH 235
 Db 179 GGSCKVTIGAEVHRIPVYQLFMPDPSRVABFNHRISGKFTYFPFFNRPNNR 238
 QY 236 LTHDCVIAPMAVALRVNVTAVAGAAHLAFDENHGAFLVLPDITTYTFQSSSSSTTTAR 295
 Db 239 LLFEAVVGAVALRCRNVDAVARAAHLAFDENHGAALPADITFTAFASQG--KTPR 296
 QY 296 GARRNDVNSTKPSGSGFERRLASIMAAADTALHAELVFNIGIYETPTDIKEWPMFQM 355
 Db 297 GGR-----DGGKGAGGFEQRLASVWAGDAALALESVMAVDFEPTDISAMPLFEGQ 351
 QY 356 EGTLPRLNALGYSYARVAGVIGAMVFPNSALYLTEVEDSGMTAKDGGPSPFNRPYQF 415
 Db 352 DTAARANAVAGAYLARAAGLVGMVFNSTGALHLETVDDAGPADPKHSHK-PSFYRFLV 410
 QY 416 AGPHLAANPQDRDGHVL-----SSOSTGSSNTEFSDVILALICGFCAPLLARLLFYL 468
 Db 411 PGTHVAANPQDRGHVVPFEGPPTAPLVGGTQ--EPAGEHLAWLCCGFPALLAKMLFYL 469
 QY 469 ERCDAGFTGGHG--DALKYVTGTDFSEIPCSICEKHTRFVCATHTVRLRORMPRFGOAT 527
 Db 470 ERCDGAVIVGQEMDVFRYVADSNQDTPCNLCFTDIEHACVHTTLMRLRAHHPKFSAA 529
 QY 528 ROPIGVGTWMSOYSDCDPLGNAPYLILRKPQGTAEAKATMODTYRATLERLFDLEQ 587
 Db 530 RGALGVGTWMSMYSDCDVLGNVAAFSALKR--ADGSETARTIMOETRYAATERVMALET 588
 QY 588 BRLLDRGAPCSSEGLSSVIVDHPTRFRLDLRLARIEQTTTQPMKVLVETRDYKIREGLS 647
 Db 589 LQYVDQAVPTANGRLTIIITREALHTVVNNVQVVDREVEQLMRNLVEGRKFKFDGLG 648
 QY 648 EATHSMALTDPYSGACFPITNFIKPKPDGTAEAKATMODTYRATLERLFDLEQ 707
 Db 649 EANHMSLTLDPYACGCPCLQLLGRSNLAVQDLALSCQCHGVFAGQSVGEGNFRNQF 708
 QY 708 PVLRRRVDFLNGGFISTRITVTTLSEG--PVSAFNPTLGDAPAGRTFDGLARVSEVI 766
 Db 709 PVLRRVDMFNNGFLBAKTLVLSGAAICAPSLTAGOTAPAESFEGDVARVTLGFP 768

185	TIGDAEVRHLPYPLQMPWPF	SRVIADPFCNCRNRSIGENFNTPLDPFNFRLARLLFEAV	244
242	IAPMAVALRVNVTAVAGAAHL	AFDENHEGAVLPDDITYTYFSQSSSGTTTARGARRD	301
245	VGPAVALARVNDVAAAAHLAF	DENHEGAALPADITFTAESQG--KPORGAR---	299
302	VNSTKPSPCGFERRRLAS	INAAOTALHAENVFNTGLYBETPTDKWPMFICMEGTLP	361
300	--DAGNKGPAQGFQRLAS	VWAGDAALALSIIVSMAVDFDPDDITWPLLEQETPPAAR	357
362	LNALGTYTARVAGVTGAMV	FSNSALYLTVEVDSGMTKADGGPFSFNRFYQFAGPHLA	421
358	AGAUGAYLARAAGLVGAMV	FTNSALHILTEVDDAGPADPKHSHK-PSFYRFFLVPGTHVA	416
422	ANPQTDGRDGHVL-----	SSQSTGSSNTRFSDYLLALICGFCAPILARLLFVLEKCDAG	474
417	ANFOLDREGHVVPYEGRP	TAPLVGGTQ-REAGHILMLCGFSPALLAMVFLERKCDGG	475
475	AFTHGHD--DALKYVTGT	SEI PCSLCEKTRPVCAHTVHRLQRMPRFGQATROPIGV	533
476	VIVGRQEMDVFRYVADSGQ	TDPVNCUFTFTRIACARHTLMRLARSHKPFASAARGAIGV	535
534	FGTWSNOSXDCDPLGNVAP	YLILKPKGDQTEAAKATMOOTYRATLERLFIDLEQERLLDR	593
536	FGTWSAYSDCDVLGNVAF	SALKAR--ADGSENRITIMQETRYATSRVMAEUALQYVDQ	594
594	GAPCSSEGLSSVVDHPT	FRRIIDTLERARIEQTTQTFMKVLYVETRDYKIREGLSEATHSM	653
595	AVPTALGRLTEIIGNREAL	TVVNNIKQLVDRVEQLMRLNLEGKFNKFRDGLAEANHAM	654
654	ALTDPYSGACPTTNFLVK	THLAVVDLALSQCHCVFGQOVGEUFRNFOFVPIERR	713
655	SUSLDPYTCGPCLLOLAR	SNLAVVDLALSQCHGVFAGOSVGRNFRNQFQVPLRRR	714
714	FVDLPNGFISTRITVTL	SEG--PVSAENPTLSQDAPAGRTFDGLARYSVSEVIRDIRVK	772
715	VMDLFNNGFLSAKTLT	VALSEGAAICAPSLTAGCTAPAESSEFGDVARVTLGFPKELRVK	774
773	NRVPSGNTNLSEAAARLV	GLASAVQCEKEVOMLHGALFLLKQFHLLFPFGMEPN	832
775	SEVLPFAGASACASAA	RVASIQSAYQKDPKKVDLILGFLGLKQPHAVIFPNKPKPG	834
833	SKSNPQNFMTLORNQ	PADKLTHEBIITIAAVKFTTEEYAAINFILPPTCIGELAQF	892
835	SNQNPQWFMTALQ	RNLPARLSREDIETIAFKRFSLDYGAINFILNAPNNVSELAWY	894
893	YMANILLYCDHSOYLNT	LTSIITCARPPROPSVLHWIKRQVTSADIDETOAKALLEK	952
895	YMANQLRYCDHSYTF	INTLAVIAGSRPPSPVQAAAWAFO--GGAGLEAGARALMDS	951
953	TENPELWTTAFTSTH	LVRAMNQRPMVVLGISIKYHGAAGNNRVFOAGNWSGLNGKN	1012
952	LDAHFGAWTSMFAS	NLLARFVMAARPMVVLGLSISKYVGMAGNDRVQAGNWAASLLGKN	1011
1013	VCPLFTDRTRBTILAC	PRGGFICPTVGSSGNRETTILSDQVRGLIIVSGGANWLAIYAT	1072
1012	ACPLLIIDRTKRVFLAC	PRAGFVCAASSIGGGAHSHSLCEQLGIIAEGGAARVSSVFVA	1071
1073	VVRVARGAQMADFND	WLSITDDEFLARDELSHDIQIITLETPTWVEGAL---EAVKI	1128
1072	TVKSLGPRTOOLQI	EDWLALLEDEYLSMEMFETTRALERGHGEMSTDAL EVAHEAEL	1131
1129	LDEXTADGGETTNLA	FNFDSCFESHETTNSVINISGSNTVGLKRPPEDELD	1188
1132	VSQUGAAGE-----	VFNFGDDEDDHRAASFGLAAA--AAGAAGVARKRAHFHGDPPG	1183
1199	LSGIPKIGNITMEM	1203	
1184	-EGPPEKK-DLTIDM	1196	

RESULT 8

DNBBEG

DNA-binding protein 5 - bovine herpesvirus 2 (strain BMV)
C:Species: bovine herpesvirus 2
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 07-Jun-1996
C:Accession: A29242
R:Hammereschmidt, W.; Conraths, F.; Mankertz, J.; Pauli, G.; Ludwig, H.; Buhk, H.J.
Virology 165, 388-495, 1988
A:Title: Conservation of a gene cluster including glycoprotein B in bovine herpesvirus 2
A:Reference number: A94381; MUID:88306231; PMID:2841793
A:Accession: A29242
A:Molecule type: DNA
A:Residues: 1-1186 <HAM>
C:Cross-references: GB:M21628
C:Superfamily: herpesvirus DNA-binding protein
C:Keywords: DNA binding

Query Match	46.6%	Score 29333	DB 1	Length 1186
Best Local Similarity	48.7%	Pred. No. 2.2e-214		
Matches 585	Conservative 188	Mismatches 373	Indels 56	Gaps 11
2y 1	MENTQKVT---VPTGPGVYACRVEDDLDEISFLAARSTDSDIALPLMRNITVEKT	57		
2b 1	MENKQKATTVKVSPPGLVYVYARLPPEGLTEALLSARSADSDTAVLPIAGLTVESG	60		
2y 58	FTSSLVAVSGARTTGLAGAGITTLKTTSHSPYPSVFVHGKHLVLPSSAPNLTACNAAR	117		
2b 61	FDVNAVVGSRITGVGGTGVSLMPSHVAPSAVYVFGHRLHAPSSAPNLSL:CDRAR	120		
2y 118	ERFGSFCQGPVVDGAVETTGASICTRLGLEPENTILYLVATLFKEAVFMCNPFHYGG	177		
2b 121	VQGFSSFKPKPCEAGETTGECALCEHLGLNPESLLYVIAEGFKEAYISNTILHGG	180		
2y 178	LDIYHINHGVDIRPLPFVQLFMPDVNRLVPDPFNTHRSIGEGFYVTPFTNGTCLHLI	237		
2b 181	VGTVTIAGEVRRIPITPLQMFMPDYCRVAVDPFNDRHSAIGEPAYFLPFPFNKASLL	240		
2y 238	HDCVIAPMAVALRVNVTAVARGAHLADENHEGAVLPDITTYTFYFSSSGTITARGA	297		
2b 241	FGAAVGAVALRARNVDVARAAHIALDENHEGALPADITTFAPDPTGS-----KAG	295		
2y 298	RENDVNSTKPSGGPERRLASIMAADTALHAENVFNTGIYEETPTDIKWPMPFIMGEG	357		
2b 296	HRNPREC-----GGGFEQRLASWAGDAALALESIMSAVPEEPTDITGWPLTCQES	349		
2y 358	TLPLNALGSYTVARVAGVIGAMVFPNSALYTEVEDSGMTEAKDGGPSPFNPYQFAG	417		
2b 350	TAAEASIGAYLGRAAGLVGAMVFPSSALHTEVDDAGPADPKD-PTKPSFYFFLVPVG	408		
2y 418	PHLAANPQTRDRGHVLLSSOS-----TGSSTNTEFSVDYLALICGFGAPLIARLLFYLERC	471		
2b 409	TYVAANPQLDRDRGVVAGHGRPIVPGVGNHEFTCEHLATL:CGFPELLAKMLYLIERC	468		
2y 472	DAGAFTG-GHGDAKLVYCTGFDPSETPCSICEKHTPVCATHTVHRLKQMRPFQCATRQP	530		
2b 469	DGGVILGRPMDTFKVVSUSAHTDPCCLLCSLDNHSCHAHTLLRLARHPKFTSTTRGA	528		
2y 531	IGVFGTMNSYSDCDPLGNVAPYLILRKPGDQTEAAKATMODTYRATLERLFDLQERL	590		
2b 529	IGIFGMVNSAYSDDVLGNVAFSAIKM-DVQETARAIMQETVRSVAERVMAELEMLNY	587		
2y 591	LDRGAPCSSEGLSSVIVDHPTRFRILDTLURARIEQTTTFQFMKVLVETRDYKIREGLSEAT	650		
2b 588	IDAAVPTSPAKLESITIGREALQTVSVNVKQVVDGEVAQLMRALVEGGRFFREALGEAN	647		
2y 651	HSMALTFDPVSGAFCPITNFWLKRTHLAVODLALSQCHCYFCQVQGVGNGRPNRQCPVL	710		
2b 648	HANSLTLDPHASVCPCLQVGLGRNLAVYQDLASQCHGVFEGQAVGEGNFRSQPQVPL	707		
2y 711	RRRFVLENGGFTSTRSITVTLSEGP-VSAPNPTLQDAPAGRTFDGDLARVSVEVIRDI	769		
2b 708	RRVLDMFNNGFLSARTLTVALTDGACISAPOLVSGQHAAAESGFCGVARVNLGPKEI	767		
2y 770	RVKQRVYVPSGNCNTI-SEARARLVGLASAYQREKRVDMHLGALGFLLLQFHGLLFFRGM	829		

```

Db      768  RVKSRVLFAGAPPAASEAARARIAGLQSAQSKDKRVDDILGLPLGLFMLKQFATLFPNGK  827
QY      830  PPNKSNPNQPFQFWTLLORNQMPADKLTHEETITIAAVKRFTEEYAAINFNLPTCTIGEL  889
Db      828  PPGSDNPNQPFQFWTALQENQLPARLLSEEDISLIAFKVRFSEYVGAGNFVNLPPNNISEL  887
QY      890  AQFYMANILIKYCHDSQYLINTLSIITGARPRDPSSVLHWIRKDVTSAADIEQAKAL  949
Db      888  ANYYNANQILIKYCDHSTYFINTLTALAGSRPPNQAQAAAAPRGGT--ELEAQARSV  944
QY      950  LEKTENLPELWTTAFTSHLVRAAMNORPMVLGISISKYHGAAGNNRVFQAGNWSGLG  1009
Db      945  VANPGDHGAMWTWFASCNLLRPWATPRPMVLGLSISKYGVAGNDRKVFQAGNLANLLG  1004
QY      1010  GKNVCPLFTFDRTRRFIIACPRGFIQVTPGSSGNRETTLSQDVRGIIVSGGAMVQLAI  1069
Db      1005  GKNACPLLIQFTRTKFYIACPRAGFVCAAVSAGSAGHESLCEQLRAI IAEAGGATVASDV  1064
QY      1070  YATVVRVARGAQAQMAFDWMLSTDDEFLARDLEELHDOIQTLETPWTVEGALEAVKIL  1129
Db      1065  FAAAKSILGARVQQLQIEDWLLALLEDEYLSSEWMLAGRALERGGGSEWSDAALDVAREA  1124
QY      1130  DEKTT-----AGDGETPTNLAFNDS-----CEPSHDTTSN  1160
Db      1125  EAMVTRHVDAAETDFGAFAEQDPADAGLAVHLQSRRLPLACSDLFCDAFAEKENDLTLD  1184
QY      1161  VL 1162
Db      1185  ML 1186

RESULT 9
JQ0846
DNA-binding protein - equine herpesvirus 1 (fragment)
C/Species: equine herpesvirus 1
C/Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 23-Feb-1997
R/Accession: JQ0846
submitted to JIPID, January 1991
A/Reference number: JQ0846
A/Accession: JQ0846
A/Molecule type: DNA
A/Residues: 1-375 <BEL>
C/Superfamily: herpesvirus DNA-binding protein
C/Keywords: DNA binding; nucleus

Query Match      17.4%; Score 1094.5; DB-2: Length 375;
Best Local Similarity 55.5%; Pred No. 3.1e-75;
Matches 212; Conservative 57; Mismatches 98; Indels 15; Gaps 5;

QY      828  GMPNKSQNPQFWTLLORNQMPADKLTHEETITIAAVKRFTEEYAAINFNLPTCTIG  887
Db      2   GHPPGIDTPQFWTLLORNQMPARLLSKEDIEITIAKRPSEDEYSAINFLTPNNIG  61
QY      888  ELAQFYMANILIKYCHDSQYLINTLSIITGARPRDPSSVLHWIRKDVTSAADIEQAK  947
Db      62  ELAQFYFANLVLIKCHDSQYFINGLTAIVGSRPRDPAVLAWIDRTINGAADYEPAAQ  121
QY      948  ALLEKTENLPELWTTAFTSHLVRAAMNORPMVLGISISKYHGAAGNNRVFQAGNWSGL  1007
Db      122  EVLQRLGSPFAAWTGFTSTNVRVMDORPMVVLGLSISKYSGAGNNRVFQAGNWNGL  181
QY      1008  NGKKNVCPLFTFDRTRRFIIACPRGGFTCPVTGSSGNRETTLSQDVRGIIVSGGAMVQL  1067
Db      182  NGKKNVCPLMAWDRTRRFVLAICPRVGFCEAGGFGTGVRENTLSEQVRGIVSEGGPMVQT  241
QY      1068  AIYATVVRVARGAQAQMAFDWMLSTDDEFLARDLEELHDOIQTLETPWTVEGALEAVK  1127
Db      242  AVFAAVLHALGARTQHLAVDDWIGLVDDFEFLAASLDALNAIYVDPQF-CENSVFAAQELVK  300
QY      1128  ILDEKTTAG---DGETPTNLAFNDS---BPSHDTTSNVLNIGSNISGTVFPLKRP  1181
Db      301  NMEAQTFAVAAGSG---APDFGACVGDTPQOSTSA----FNGGLMAAAPAGQKESL  352

```

QY 1182 EDELFDLGSGPIKHGNTM 1203
 Db 353 PDDLFDGAPPEKSGLTFFDM 374

RESULT 10
 QOE47
 DNA-binding protein - human herpesvirus 4 (strain B95-8)
 C:Species: human herpesvirus 4, Epstein-Barr virus
 C:Date: 25-Feb-1995 #sequence_revision 25-Feb-1995 #text_change 16-Jul-1999
 C:Accession: A03045; A03791; S33057
 R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
 Mol. Biol. Med. 1, 21-45, 1983
 A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
 A:Reference number: A93065; MUID:85035713; PMID:6082825
 A:Accession: A03045
 A:Molecule type: DNA
 A:Residues: 1-1128 <BAN>
 A:Cross-references: EMBL:V01555; NID:G59074; PIDN:CAA24808.1; PID:G1334916
 R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
 Nature 310, 207-211, 1984
 A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
 A:Reference number: A03794; MUID:84270667; PMID:6087149
 A:Contents: annotation; protein coding region
 C:Superfamily: herpesvirus DNA-binding protein
 C:Keywords: DNA binding

Query Match 10.6%; Score 669; DB 1; Length 1128;
 Best Local Similarity 23.2%; Pred. No. 4.8e-42;
 Matches 292; Conservative 186; Mismatches 537; Indels 246; Gaps 45;

2Y 13 GPLGYVACVDELDLEISFLAARSTDSALLPLMRNLTKETTSLLAVVSGARTG 72
 Db 18 GPGCIYFYPLATYPLREVATLCTGVAGHRCITVPLLCITVEPGFSIN---VKALHRRP 74
 2Y 73 LAGAGITLKTTHSHFYPSVFHGGKHVLPSSNAPNLTACNAARPFQFS-----RCQ 126
 Db 75 DPNCGI---LRATSHRDYVFHNAHVPPIFEGPGLCALCGETREVFYDAYSALPRES 131
 2Y 127 GPVVDGAVETGAEICTRLGLBEPNTILYLVATLAFKEAVFNCVFLHYGGLDIHINHG 186
 Db 132 SKGDPFPE-----GLDPSAVLGAVATEAFKELYSGNLVAPSLKQEVAVGQS 181
 2Y 187 DVIRIPFPVQVMPVNVNBLVDDPNTHRSIGEGVPVTPFYNTGLCHLIHDCVIAPMA 246
 Db 182 ASVRVPLDYKEVPFEGVPQI-----RQFNSDLKRCMHEALYTGLA 222
 2Y 247 VALRVNNTAVARGAAHLAFDENHGAVALPPDITYTFQSSSGSTTARGARRNDVNSTS 306
 Db 223 QALRVARVGL-----VELLEKQSLQDQAKVAKVAPLKEFPAST 261
 2Y 307 KSPSGGPFERRLASMAADT-ALHAEIFNTGIYE---ETPTDIK--EWPMPFIMEGTLP 360
 Db 262 ISHPDSG-----ALMIVDSAAELAVSYAPAMLEASHETPASLNYDSWPLFADCEGPEA 315
 2Y 361 RNLALGSYTARVAGVIGAVFSPNGLALYTEVDSGMTAEKDGPGSPFNFPYAG--- 417
 Db 316 RVAAHRYNVAAPHVSTQIFATNSVLVY-----SGVSKSTGQCKESLFSFYTHGLGT 370
 2Y 418 -----PHLAANPQTDRGDHVLSSQSTGSSNTSFSDYALICGFCAPILARLLF 466
 Db 371 LOEGTWDPCRRPCFSGWGPDTG-----INGPQN--YAVEHLVYAASFNLLARYAY 422
 2Y 467 YLERCDAGFTGGHGDAL-----KYVTGTFDEIPSLCEKHTRPVCAHTTVHRLQR 519
 Db 423 YLQFCQ-----GQKSLTVPVETGSYAGAAAFPM-CSLCEGAPAVCLNTLFFRLDR 475
 2Y 520 MPRFQATQRTGIVFGTWNVSQDCDPLNVPAYLIL---RKPGDQTEAAKATMQTYR 575
 Db 476 FPPVMSVTRQRPYVTSAGSGSYNETDFLGNFLNFDKEDDQCRDDEPRYTYWLNQNL 535
 2Y 576 ATLERLFDLQERLLDRGAPCSSEGLSSVIVDHPTRRLDILRLARIEQTTQFMKVLV 635

Db 536 ERLSLGIDAECKLEKEPHGP-----RDFVKMFQVDAAVDAEAVQFMASMA 582
 QY 636 ETDRYKIREGLSEATHSNALTDFPYSGAFCDITNFWLVRTHLAVQDLALSQCHCVYQG 695
 Db 583 K-NNTYKDLVRKSCYHVMOYSCNPFQAPACPIFTQLFYRSLTLIQDISLFCMC----- 636
 QY 696 QVEGRNFRNQFVLRRRFVDFLNGGF-----ISVRSITV-----TLSEGPVSAFNPITLG- 745
 Db 637 -----YEND-NPGLQGSPPEMLKGHVQTLCTNFRSLAIDKGVLTAKKVVHGEPTCDL 689
 QY 746 -----QDAPAGRTFDGLARVSEVIRDIRKRVVFGNCTNLSEAAARLVGLASA 798
 Db 690 PDLDAALQGRVYGRRLPVRMSKVLMLCPRNKIKRNVVFTGENAALQNS-----PIKS 742
 QY 799 YQREKRVDMHLRGALGFLKQFHLLPFRGMPNPKSPNQWFLLQRMQPADK-LTH 857
 Db 743 TTRRENYI--INGPYMKFLNTYHKTLP-----PDKLSSLYLWHNFSRRSRVPSGASA 795
 QY 858 EBITTIAA-VKRFTEEVAANFINLPTTCIGELAQFYMANLILKYCDHSQYLINTLTSII 916
 Db 796 EYSDLALFVDGGSGRAHEESNVIDVPGNLVTVAKORLNNAILKACQQTQFYISLIQGLV 855
 QY 917 --TGARRPRDPSVVLHWRKDVTSAAADIETQAKALLEKTENLPELWTTAFTSTHLVRAAM 974
 Db 856 PRTQSVFARDYPHVLG--TRAVESAA-----AYAEATSSLTAT-TVVCAATDCLSQVC 905
 QY 975 NORPMVVLGISIKYHGAAGNRVFOAGNWSGLNGKNVCPLFTFDRTRRFFIACPRGG- 1033
 Db 906 KARDVWTLPTVINKYTGWNGNNOIFQAGN-LGYFMGRGV-----DRN---LLOAPGAGL 955
 QY 1034 -----FICPTVGPSSGNR-ETLSDQVRGIVSGGAMV-----QLAIYATV 1073
 Db 956 RKQAGSGSMKKVFATPTGLTVKERTQAATTYEIENRAGLEAIIQKQEBDCVFDVV 1015
 QY 1074 VRVAGARQAHFADDMLSLT--DDEFLARDLELHQIITQLET-----PMTVEGALEA 1125
 Db 1016 CNLVDAMGEACA-----SLTRDDAEVLLGRFVSLADSVLETATASSIETAEARDF 1070
 QY 1126 VKILDEKTTAGDGETTNLAFNFDSCPSHDHTTSNVLNTSGSNISG-----STVPGL 1177
 Db 1071 L-----EGWVGPGGAQDNFISVAEPYSTASQASAGILLGGGGGSGGRKRRLATVPLSL 1126
 QY 1178 K 1178
 Db 1127 E 1127

RESULT 11
 T42922
 major single-stranded DNA binding protein - ateline herpesvirus 3 (strain 73)
 C:Species: ateline herpesvirus 3
 A:Variety: strain 73
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 19-Feb-2000
 C:Accession: T42922
 R:Albrecht, J.C.; Fleckenstein, B.
 submitted to the EMBL Data Library, August 1998
 A:Description: Primary structure of the herpesvirus ateles genome.
 A:Reference number: Z22274
 A:Accession: T42922
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1128 <ALB>
 A:Cross-references: EMBL:AF083424; PIDN:AAC95531.1
 A:Experimental source: strain 73
 C:Superfamily: herpesvirus DNA-binding protein

Query Match 10.1%; Score 633; DB 2; Length 1128;
 Best Local Similarity 22.4%; Pred. No. 2.6e-39;
 Matches 274; Conservative 194; Mismatches 514; Indels 240; Gaps 41;

QY 11 PTGPGYVACVDELDLEISFLAARSTDSALLPLMRNLTKETTSLLAVVSGART 70

18 PVGAGYIYVPEGPPFQEQASLIGNNNGADAVSLLSGLTVEANFENFNVKAVHKID 77
 71 TGLAGGITLKTSHFPYVFFHGGKHVLPSSAAPNLTRACNAARERFGFSRCQGPV 130
 78 M-----TTLVRSVAVHREAIYFNTNFTPIFFPGGLDLSLSDARNLFGVTSFT-PRH 130
 131 DGA-----VETGAEICTRLGLEBENTILYLVVLTALKEAVFMCMVFLHYGGLDIVHNG 186
 131 DLKDLIDIKLYAPFYTK-----DSCFMSVVVTFESFERLYFNGNLVPLISQOKVQINGR 185
 187 DVIRIPLFPVQLEMPDVNRLVPDPFNTHRSIGEGFVYPTFPNTGLCHLIHDCVIAPMA 246
 186 EAVKIPLYDEDFSKSHEHL-----PRFIPSVKYLHDSLFTSTA 226
 247 VALRVNNTAVAGAHLAFDENHEGAVLPDPITYYFQSSSGSTTTARGARRDNVNSTS 306
 227 QALRIKRVESVIRAEKQSTHDOYKLAKV-----VNSK- 259
 307 KPSPGSGFFERRIASMAADA-----LHAEVIFNTGI-YEETP-----TDIKENWMP 352
 260 -----EFALQAVKQDASAPFVWIDCTAAELAISYGLSFLFTQEPCCALLDYTSNPIF 311
 353 IGMEGLPLNALGYSYARVAGVIGAMVSPNSALYLTVEGSMTEAKDGGGPGSPENRF 412
 312 DTTEBEGRIKAIQDNWAMSVHVVTHLFTSTNSVLYLKINKQTOFNKSDQV-----IYNSY 368
 413 YQFAGPHLANPQTDGHLVLSQSTGSSNTBFSVDYLALICGFGAPLLARLLFYLERCD 472
 369 FMQGLSYAAEAATQKENGDPAPFSAVKFNGGYSYTLFHLALASSFPHLLARNCYMQPCQ 428
 473 AGAFTGGHGDALKYVTGTFDSBIPCSLCEKHTRPVCAHTVHRLRQRPGRQATRPQIG 532
 429 HQKSTNSYSAVQVGVIVAVSELCELCQKCPAACIHTLFRUKDRPPVLLSQRRDPY 488
 533 VEGTWNYSQDCDPLGNVAPYILILKPGD-----QTEAAKATMQDTYATLRLPIDLEQE 588
 489 VVTGSGQYNDLMLGNFATF-----REKEDDMVQNTCEKYTWGLIQNVVEKL----- 538
 589 RLLDRGAPCSSEGL--SSVIVDHPFTRILDTLRLARIEQTTQPMKVLVETRDYKIREGL 646
 539 -----ATIGITEGLGSLIITDIQSLKTFREIDNVVDNEVIRKINCLVK-NKINFRETI 592
 647 SEATHSMALTDPYSGAFCPITNELVLRKTHLAVQDLALSCHQCVFYGO----- 695
 593 KGVHVLHYCCNVFQWQPCAMFLNFYKSVLAIQDCLPV--AMIVEQDNPSMGMPSE 650
 696 -----QVSGNERN-----QFQVLRFRFVDLFGNGFISTRSITVLTSEGPVS 738
 651 WLMHYQITWTFKSSCIDKGLTGASHKVVHGMFCDFINVD-----S 694
 739 APNPTLGDAPAGRTFGDLARVSEVIRDIRVKNRVVFSNGCTNLSEARARLVGLASA 798
 695 ALN--GQIVPV--KMQRVLAKALITVPKTIKIKRIVFEN--SSMTEATQAGFI----- 742
 799 YQREKRVQD--MLHGALGFLKQPHGLLFRGMPNPSKSPNQWF--TLQRMQMPA-DK 854
 743 --RSSTKDSYIVTGYMKFLNSLHKVIFP-----DAKISALYLVHWTFSQKEQIPVLP 794
 855 LTBEIETITIA-AVKRETEEYAAFINPLPTCIGELAQFWMANILIKYCDHSYLNITL- 912
 795 ISKENIITELANYIEAGSMHDDMMVLDIIPITLAVAKVELANNTILRTCGQTOFYATTLQ 854
 913 -----TSIITGARRPRDPSSVLHWIRKVDTSAADETQAK-----ALLEKTENIPELMTTAF 964
 855 CLLPITLQWASATEYH-----VLH--QOSITSVDVYLSSIKNKQALLIVQTLKEDIATIG- 907
 965 TSHLVZAAANQBMVVLGISIKYHGAAGNNRVFOAGN-----MSGNGNGNVCPLETF-ED 1020
 908 -----KHRPIVTVPLVNVKTYGINTQITFOCGNLGFWMGVGRVDRNLIPSSGGR 957
 1021 RTRRFIIACRGGGFIQVTC-----PSSGNREI-----TSLDQVRGIIVSGGAM 1064
 958 RONNSAYMRKHFIMTPIVANLKRISNLNLTFFVETIRKKNVQITFDKDNLIPDNVV 1017

QY 1065 VOLAIYATVTVAVGARQAHVAFDWMLSLTDREFLARDEELHQIITQLFTPTWTEGAL 1124
 Db 1018 IEL-----VKLGESCENTEDDLQFYLGDIYIMDEIWSRFQLLTDSGTSPWSVE--S 1068
 QY 1125 AVKILDEKTTAGDGETTNLAF 1146
 Db 1069 VTILGSRK-----QEDCNLEF 1085

RESULT 12
 SS5600
 single-stranded DNA binding protein 06 - equine herpesvirus 2
 C/Species: equine herpesvirus 2
 C/Date: 27-Oct-1995 #sequence revision 03-Nov-1995 #text change 26-Aug-1999
 C/Accession: S55600
 R/RefSeq: E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.
 J. Mol. Biol. 249, 520-528, 1995
 A/Title: The DNA sequence of equine herpesvirus 2
 A/Reference number: S55594; MUID:95302501; PMID:7783207
 A/Accession: S55600
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-1145 <REL>
 A/Cross-references: GB:U02824; NID:9695172; PIDN:AACI3793.1; PID:9695178
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995
 C/Superfamily: herpesvirus DNA-binding protein

Query Match 9.8%; Score 618; DB 2; Length 1145;
 Best Local Similarity 23.1%; Pred. No. 3-8e-38;
 Matches 295; Conservative 182; Mismatches 540; Indels 260; Gaps 47;

QY 4 TQKTVPTGPIGYYVACRVEDLLEESFSAARSTDSDLALILPLMRNLTVKFTSSIA 63
 Db 17 TQASV---GPGFVLYIPGDTFPVEASLLGNLHAGVEFSLPLSLGLTVEADPHNVK 72
 QY 64 VVSGARTTGLAGITLKLTTSHFYPSVFVHGGKHVLPSSAAPNLTRACNAARERFGFS 123
 Db 73-AVHKILDEP-----TVSKASAYHREIVFANAACFKPIGAPGLEGLCAASQLFGYA 126
 QY 124 RCGGPPVDGAVETTCG-----EICTRLGLEP--ENTILYLVLTALKEAVFMCMVFLHYG 176
 Db 127 EPE-----BRAGGAARPFELADLHLLGASHIAGVYVTFESFERLYRGQLVAVVES 178
 QY 177 GUDIYVHNGVDVIRLFPFVQVLFMPDVNRLVPDPFNTHRSIGEGFVYPTFPNTGLCHL 236
 Db 179 QIOSVRVGECDFAFKVPLVDGELFAKSPCR-----ENLRY---FYHAGVSRV 221
 QY 237 THDCVIAPMAVALRVNNTAVARGAHLAFDENHEGAVLPDPITYYFQSSSGSTTTARG 296
 Db 222 LFEAHYVTLAQALRVDPVPGILGALERQSF---HDQVKLP-----KVYCEKPEPATGHRG 273
 QY 297 ARNDVNSTSKPSPGGFERELASIMAADTALHAFLVFNLTGIYEETP-TDIKEMPFIQM 355
 Db 274 A-----GDCSLTIVDSVATELAVSGLTFLEVPOEGTALLSYDKWPIPEGC 319
 QY 356 EGTIPLRLNALGYSYARVAGVIGAMVSPNSALYLTVEGSMTEAKDGGGPGSPENRYQF 415
 Db 320 ETPEORVEALTQFNKAQAVHVSQLFSGNSVLYLARKVQKQASN--RGGGGVNYSNPFMG 377
 QY 416 AGPHLANPQTDROG-----HVLSSQSTGSSNTBFSVDYLALICGFGAPLLARLLFYLER 470
 Db 378 HGLACLAPTKQENGLPSFPFGVPASALSGSN---YSLHLAYAASFQPMARHCYVLOF 434
 QY 471 CDAGFTGGHGDALKYVTGTFDSBIPCSLCEKHTRPVCAHTVHRLRQRPGR-FGQATQ 529
 Db 435 AQHKQSSNNSSVNVPTVVGTAANTPMLCELCGSCPASCNVNLFYRLDRPFPVAVSRD 494
 QY 530 PIGVFGTNSQYSCDCLPGLNYPVYLIRKFGDQTEAAKATMQDTYATLRLPIDLEQER 589
 Db 495 FYVVTGVAGA-YNLDLIDAGNFANY---RDXDESSNQSSEEREKFTYQVQTQVL-----ER 545
 QY 590 LLDRGAPCSSEGLSSVIVDHPFTRILDTLRLARIEQTTQFPMKVLVETRDYKIREGLSEA 649

```

Db 546 LSEAGICEGEGDVGDAIHNIIGSLFKVFEIDGIVGEVARFINSVK-NVNYRESIKSI 604
QY 650 TISMALTEDPYSGACPIINFLVKEHLAVVQDLALSOCHVFGQ----- 695
Db 605 HIIQVQVNVYQPCPFVFLNLYRCVLAIVQDCLPT--CMYEQENPAVGVSPGEWLK 662
QY 696 ---QVEGRNFRN-----QPQVLRFRFVDFEN-----GGFISTRSITVTLSSEG 735
Db 663 MHYQTLWTFNFKNSCIDKGLVITGTEYKVVHKQFCDFDVSADARGEVFCVKT-QVRLSRA 721
QY 736 PVSAPNPLTGODAPAGRTFDGLARVSVEVIRDIRVQRVVFSGNCTNLSAARLVL 795
Db 722 LMNP-----RVMKIKRIIFSNPCTES-----I 746
QY 796 ASAYQORQRVD--MLHGALGFLKQFGLLPPRGMPNPSKPNQFW-TLLQRNOMPA 852
Db 747 QNAFVRGTFPKGDSVSGPYMRFSLTHSGLFP-----GAKISFLFWHTFSGKQLPV 800
QY 853 -DKLTHEITIA-AVKRFTTEYAAINFILPPTCIGELAQFYMANILKYCDHSQY--- 907
Db 801 FPNVPRESVTELANYVEQNSRLHGETSIIDVVPENFYTYAKVRLNNALFRACGQTQFYAT 860
QY 908 LINTLITSIITGARRPROPSSVYLHWRKDV-SAADIETQAKALLEKTENLPELWTTAFTST 967
Db 861 TIHCLTPIKIQTVPAREVPHAL-----GARGVADVGEYLGAAARELT--VP---TVQCTSR 909
QY 968 HLVRANMQRMMVLGISISKYHGAAGNNRVFQAGN---MSGNGGNNVCPVLPFTFDRTER 1024
Db 910 DNICEVGKCRPIVTLPLVANKYTGVTGNSQIFOCANLGYFGRVQKULP----- 960
QY 1025 FIACPRGGGICPVTPSGSGNRE-----TTLSDVRGIIVGGAMVOLAIYATVVRVAGA 1079
Db 961 -----DAGSF--KKQGVSTMRKHVFWMTPLSDHLLRRSVQGAHV-----APEIGVRR 1007
QY 1080 RAQ-----HNAFDDWLSLT-----DDEF-----LAEDLEELHDOII 1110
Db 1008 RVQOILSDGNPHVIRDVWLQVLSGSECRSVSEYDLEVMQYVIFAGDVSE-RLQEL 1066
QY 1111 QTLTETPTVSEALEAVKILDEKTAGGETPTNLAFNFSCE-----PSHDTTNSVLN 1163
Db 1067 SDLGDMSEWALSUL-----GEEEDPLGGELEFEKVEDAECLGHPQDQDEFALAPQ 1117
QY 1164 IGSNTISG-STVPGLKR 1179
Db 1118 AAAPQVSGSSVAGKKR 1134

RESULT 13
DNBEMI
DNA-binding protein - saimiriine herpesvirus 1 (strain 11)
;Species: saimiriine herpesvirus 1
;Note: host Saimiri sciureus (common squirrel monkey)
;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
;Accession: G36806
;Albrecht, J.
submitted to the EMBL Data Library, January 1992
;Description: Primary structure of the herpesvirus saimiri genome.
;Reference number: A36806
;Accession: G36806
;Molecule type: DNA
;Residues: 1-1128 <ALB>
;Cross-references: GB:X64346; NID:960320; PIDN:CAA45629.1; PID:G60327
;Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Biesinger, B.; Newman, C.; W
J. Virol. 66, 5047-5058, 1992
;Title: Primary structure of the herpesvirus saimiri genome.
;Reference number: A37309; UID:92333688; PMID:1321287
;Contents: annotation; possible protein-coding frames
;Note: neither amino acid nor nucleotide sequence is given
;Genetics:
;Gene: 6
;Superfamily: herpesvirus DNA-binding protein
;Keywords: DNA binding

```

```

Query Match 9.6%; Score 604; DB 1; Length 1128;
Best Local Similarity 22.9%; Pred. No. 4 3e-37;
Matches 275; Conservative 189; Mismatches 509; Indels 226; Gaps 40;

QY 11 PTCPLGVYVACRVEDLDLEISIFLAARSTDSALPLMRLMTVEKTTSSLVVSGART 70
Db 18 FVPCGYIYVYKPEGFPFKASLLGNKNGVAGASMLPLSLDITVESNPFNVKAVHKID 77
QY 71 TGLAGAGITLKLITTSHFYPSVFVFGKHVLPSSAAPNLTRACNAAREFRGFSRCQGPV 130
Db 78 M-----TTLVVRVAYHREAVIFENTDFEPIFVGPGDLILCSDAKSLFGYTFV 129
QY 131 DGAVETTGAEICTRGLPEPENTILYLVVTFALFKAEVEMCN-VFLHYGGLDVIHNGEVI 189
Db 130 TDLRDTVDIKOLYAPFYSEDSCFMAVVVTEGKERLYFGLNLPVIAOGLK-VQINGREAV 188
QY 190 RIPLFPVOLFMVDVNLVDPDPFNTHRSIGSGFYVFPFNTGLCHLHDCVIAPIAVAL 249
Db 189 KIPLVDEDLFSKSEHL-----PRFYIPSVSKVLHDSVFTSIAQAL 229
QY 250 RVNVTAVARGAHLAFDENHEGAVLPDITTYTFQSSSGSTTTARGARRNDVNSTKPS 309
Db 230 KIRDVESVIRASEKSIQDQYKLAKI-----YNSK----- 259
QY 310 PSGGFERELASIMAADTA-----LHAEIFNTGI-YEETP-----TDIKWPMPIGM 355
Db 260 -----DFSLOQVKQDASAFNVIDCIAEELAMSVGLSFLFLEAPQDPCAVLDYTSNP 314
QY 356 EGTPLRNALGSYIARVAGVIGAMVFPNSALYIETVEDSGMTEAKDGGPGSPNFRFYQF 415
Db 315 ETEEDRIKAIQDMVAMSVHVTYHLFTSTNSVLYLTINKQTSQKSEON---VYNYFMQ 371
QY 416 AGPHLAANPQDRGHVLSQSGTSSNTFSDVYALICGFGAPLARLLFYLERCDAGA 475
Db 372 HGLAYADATQRENGSPAFSGAPKFGSGTTLVHLALASSFPHLLARNYVQFQHQK 431
QY 476 FTGGHGDALKVTITTFDSEIPCSICEKHTRPVCAHTVHLRORMPR-FQATRFQIGVF 534
Db 432 STTNANVSFQYVGTAAASDLCELCOGTCPCASCHILFVRLKDRFPVLSGQRDRPVVT 491
QY 535 GTMNSQVSDCDPLGNVAPYLIIRKPGDQTAAKATWDTYRATLERFLDLQERLLDRG 594
Db 492 G-VSGQYNDLDMIGNPATF---REKEDAVQNAESEKTYWQLIQNV-----EKLSTWG 542
QY 595 APCSEGLSVIVDHPPTFRILDTLRARIEQTTQFMKYLVEVDYKIEGGLSEATHSA 654
Db 543 VTSGTVG-SBLITDQSFLEKTFDINNVDSSEVSEVFMNCLVK-NNINFRFTIKTVHVLH 600
QY 655 LTPDPYSGACPIINFLVKRTHLAVVQDLAL-----SQCHCVFYGOQV 697
Db 601 YCCNVFMQAPCAMFLNLFYKSLVLAIDICLPIAMTYEQDNFSGIMWMPSEWLKVHY-QT 658
QY 698 EGRNFR-----NOFQVLRFRFVDFLNGGFISTRSITVTLSGSPVAPNPTLQ 746
Db 659 IWNPFKSSCLDRGLVTCSEHKIVHTDMFCDFLN-----IDSALS-----GQ 699
QY 747 DAPAGRTFDGLARVSVEVIRDIRVQRVVFSGNCTNLSAARLVLGLASAYQOEKRV 806
Db 700 IVP--MKQVRLAKALLTVPKTIKIKNRIVFSN--SSMTETIQSGFIKSAT-----KDS 750
QY 807 DMLHGALGELLKQFGLLPPRGMPNPSKPNQFWTLLQRNOMPA-DKLTHEITIA- 864
Db 751 YITVGYMKFLNSLHKWMP-----PNAKISALYLVHFTSQKQLFVLPGISRENNVELAN 805
QY 865 AVKRFTTEYAAINFILPPTCIGELAQFYMANILKYCDHSOYLINTLSI-----ITG 918
Db 806 YVFTSKMDDMMVLDIPTLLTYAKVRLNNTILRTCGQTQFYATTLQCLPTLQTTISA 865
QY 919 ARPR--DPS--SVLHWIR--KD-----VTSAADTETQAKALLEKTENLPELWTTAF 964
Db 866 TEYPHVLLDOSVSDHVLSSIKDKHALTVQTTLKEDIATVGK----- 908

```


889 LAQFYNANILKYCDHSQYL--INTLTS-IITG-----ARRPRDPS--SVLHWIRK 934

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 30, 2004, 13:06:32 ; Search time 47 Seconds
(without alignments)
4062.725 Million cell updates/sec

Title: US-09-769-699-2

Perfect score: 6294

Sequence: 1 MENTQKTVTPGPIGVYA.....DELFDLGIPIKGNITM 1203

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 15872573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A Geneseq 19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	6294	100.0	1203	AA252633	Varicella-Zoster v
2	3054	48.5	1452	AA225598	Herpesvirus ICP8-G
3	3054	48.5	1452	ABG73090	Herpes simplex vir
4	3016.5	47.9	1196	AAW72105	HSV-2 strain SB5
5	1566.5	24.9	623	AAW72197	HSV-2 strain SB5 C
6	1380.5	21.9	543	AAW72196	HSV-2 strain SB5 C
7	620	9.9	1132	AA253126	Macaca mulatta rha
8	183	2.9	35	AA226935	Varicella zoster v
9	183	2.9	35	AA273931	Varicella zoster v

10	156.5	2.5	1841	21	AA07562	Protein encoded by
11	125.5	2.0	19938	22	ABP76682	Streptomyces virid
12	123.5	2.0	3210	24	ABP73958	Myxelia sterilia c
13	123	2.0	1074	20	AAV00206	Enterococcus faeca
14	123	2.0	1074	20	AAV00188	Enterococcus faeca
15	123	2.0	1074	23	ABF43407	Enterococcus faeca
16	123	2.0	1074	23	ABP43425	E faecalis EF094 p
17	123	2.0	1074	24	ABU13686	E faecalis EF102 p
18	123	2.0	1074	24	ABU13704	Enterococcus faeca
19	122.5	1.9	901	22	ABF58224	Enterococcus faeca
20	122.5	1.9	2387	21	AAV53655	Drosophila melanog
21	121	1.9	1069	21	AA841326	Mechanical stress
22	121	1.9	3158	23	ABF62763	Human ORFX ORP1700
23	120.5	1.9	2597	21	AAV53664	S. roseosporus dap
24	120.5	1.9	2597	23	ABG32889	Mechanical stress
25	120.5	1.9	2597	23	ABG32890	Rat osteoclast pro
26	120.5	1.9	2597	23	ABG31316	Rat osteoclast pro
27	120.5	1.9	2597	23	ABG47933	Rat protein 508.
28	118.5	1.9	1076	22	ABG20205	Rat OCP. Ratus r
29	118	1.9	1784	22	ABG24772	Novel human diagno
30	115	1.8	1150	14	AAW43950	Ge protein fragmen
31	114.5	1.8	1496	20	AAW81030	Melanoma associate
32	114.5	1.8	1496	21	AAV70469	Human p53 target m
33	114.5	1.8	1496	24	ABU03498	Angiogenesis-assoc
34	114.5	1.8	1498	22	ABB11587	Human peroxidasin
35	114	1.8	774	7	ABP60616	Cephalosporin C ac
36	111	1.8	4530	18	AAW19629	Streptomyces venez
37	111	1.8	4530	21	AAV77177	S. venezuelae vep
38	110.5	1.8	1532	22	ABP68585	Drosophila melanog
39	110.5	1.8	3070	22	AAW81195	Mycobacterium tube
40	109.5	1.7	2630	24	ABG76186	Human serine/threo
41	109.5	1.7	7968	24	ABG76187	Human serine/threo
42	109	1.7	11096	22	AAE10129	Streptomyces nous
43	108.5	1.7	1451	22	AAE95868	Human protein sequ
44	108	1.7	2596	22	AAW30569	A splice variant o
45	106.5	1.7	774	16	AAW86374	Cephalosporin C ac

ALIGNMENTS

RESULT 1	AAW52633	standard; Protein, 1203 AA.
ID	AAW52633	standard; Protein, 1203 AA.
XX	AAW52633	
AC	AAW52633	
XX	AAW52633	
DT	18-FEB-2002	(first entry)
XX	Varicella-Zoster virus (VZV) ORF29p protein.	
DE	Varicella-Zoster virus (VZV) ORF29p protein.	
XX	ORF29p; open reading frame 29p protein; Varicella-Zoster virus; VZV;	
KW	drug delivery; gene delivery; fusion protein; protein secretion.	
XX	Human herpesvirus 3.	
OS	Human herpesvirus 3.	
XX	WO200154709-A1.	
PN	WO200154709-A1.	
XX	02-AUG-2001.	
PD	02-AUG-2001.	
XX	25-JAN-2001; 2001WO-US02500.	
PF	25-JAN-2001; 2001WO-US02500.	
XX	25-JAN-2001; 2000US-0177901.	
PR	25-JAN-2001; 2000US-0177901.	
XX	(UYCO) UNIV COLUMBIA NEW YORK.	
PA	(UYCO) UNIV COLUMBIA NEW YORK.	
XX	Silverstein S, Annunziato P, Gershon A, Lungu O;	
PI	Silverstein S, Annunziato P, Gershon A, Lungu O;	
XX	WPI; 2001-488744/53.	
DR	WPI; 2001-488744/53.	
XX	N-PSDB; ABA02175.	
DR	N-PSDB; ABA02175.	
XX	Composition of matter for delivering desired compounds into eukaryotic	
PT	cell, comprises Varicella-Zoster virus open reading frame 29p protein	

09/769,699
2-21-04
Search Notes

either bound to an agent or operably affixed to lipid-soluble group -

Claim 1; Fig 6; 68pp; English.

This sequence represents the Varicella-Zoster virus (VZV) open reading frame 29p (ORF29p) protein. Although ORF29p is the major DNA binding protein of VZV, the invention is based on the discovery that the VZV ORF29p protein can readily enter and exit eukaryotic cells. Accordingly, the invention relates to a composition for delivery of an agent into a eukaryotic cell, comprising the VZV ORF29p protein bound to a polynucleotide, or an organic compound. The invention also relates a composition in which the ORF29p protein is linked to a lipid soluble group that permits the protein to be anchored to a lipid membrane, and lipid vesicles containing this composition. The invention additionally relates to the use of the ORF29p protein as a facilitator of secretion of a desired protein, in which the desired protein and the ORF29p protein are recombinantly expressed as a fusion protein. The invention further encompasses nucleic acids encoding the VZV ORF29p protein and their use in detection of VZV ORF29p nucleic acids, and a monoclonal antibody against the ORF29p protein. Compositions of the invention are useful for the delivery of a prophylactic or therapeutic agent to a eukaryotic cell, particularly a human cell. Examples of agents that can be delivered to a cell include proteins such as insulin, factor VIII, factor IX, and procereases; polynucleotides (e.g., for use in gene therapy), and organic compounds such as vitamins and a wide variety of pharmaceuticals (e.g., antineoplastic, antidiabetic or immunosuppressive drugs).

Sequence 1203 AA;

Query Match 100.0%; Score 6294; DB 22; Length 1203;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1203; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MENTQKTVTPTGVLGVYACRVEDDLLEISFLAARSTDSALPLMNLVETKFTFS 60
 1 MENTQKTVTPTGVLGVYACRVEDDLLEISFLAARSTDSALPLMNLVETKFTFS 60

61 SLAVSGARTTGLAGITLKLTTSHFYSPVVFPHGKHVLPSSAAPNLTRACNAAREF 120
 61 SLAVSGARTTGLAGITLKLTTSHFYSPVVFPHGKHVLPSSAAPNLTRACNAAREF 120

121 GFSCQCPVVDGAVETTGAEICITLGLPEPNTILYLVTALFKEAVFMCNVLHYGGLDI 180
 121 GFSCQCPVVDGAVETTGAEICITLGLPEPNTILYLVTALFKEAVFMCNVLHYGGLDI 180

181 VHINHGDIIRIPFPVQLFMDVNLVPDPFNTHRSIGEGFYVPTFYNTGLCHLHDC 240
 181 VHINHGDIIRIPFPVQLFMDVNLVPDPFNTHRSIGEGFYVPTFYNTGLCHLHDC 240

241 VIAPMAVALVRNVAVARGAAHLAFDENHGAVALPDPDITYTFQSSSGTTARGARN 300
 241 VIAPMAVALVRNVAVARGAAHLAFDENHGAVALPDPDITYTFQSSSGTTARGARN 300

301 DVNSTKSPSPGGERLRLASIMAADTALHAENVFNTGIEEPTDKEWPMFIGNETLP 360
 301 DVNSTKSPSPGGERLRLASIMAADTALHAENVFNTGIEEPTDKEWPMFIGNETLP 360

361 RLNALGSYARVAGVIGAMVFPNSALYLTEVEDSGMTEAKDGGPSPFNRYQFAGPHL 420
 361 RLNALGSYARVAGVIGAMVFPNSALYLTEVEDSGMTEAKDGGPSPFNRYQFAGPHL 420

421 AANPQTRDGHVLSSTGSSNTFSVDYALICGFGAPLLARLLFYLERCDAGFTGGH 480
 421 AANPQTRDGHVLSSTGSSNTFSVDYALICGFGAPLLARLLFYLERCDAGFTGGH 480

481 GDALKVVTGTFDSEIPCSICEKHTRPVCAHTVHLRQMRPFGQATQPIGVFTWNSQ 540
 481 GDALKVVTGTFDSEIPCSICEKHTRPVCAHTVHLRQMRPFGQATQPIGVFTWNSQ 540

541 YSDCDPLGNVAVYLILRKPDDQTEAAKATMDTYRATLERLIDLEQRLLDRGAPCSSE 600
 541 YSDCDPLGNVAVYLILRKPDDQTEAAKATMDTYRATLERLIDLEQRLLDRGAPCSSE 600

601 GLSSVVDHPTFRRIIDTLRARIQITTOFMKVLVETEDYKIREGISEATHSMALTEDPY 660
 601 GLSSVVDHPTFRRIIDTLRARIQITTOFMKVLVETEDYKIREGISEATHSMALTEDPY 660

661 SGAFCDITNFKVTRHLAVVDLALSQCHCVFYGOQVEGRNFRNQFQVLRFRFVDFLNG 720
 661 SGAFCDITNFKVTRHLAVVDLALSQCHCVFYGOQVEGRNFRNQFQVLRFRFVDFLNG 720

721 GFSTSTSIIVTISEGPVSAPNLTGODAPAGTFFDGLARVSEVIRDIRVKNRVVFSGN 780
 721 GFSTSTSIIVTISEGPVSAPNLTGODAPAGTFFDGLARVSEVIRDIRVKNRVVFSGN 780

781 CTNLSEARARLVGLASAYQREKEVDMLHGAIGLFLKQFHGLLFPFGMPSPNSKPNQW 840
 781 CTNLSEARARLVGLASAYQREKEVDMLHGAIGLFLKQFHGLLFPFGMPSPNSKPNQW 840

841 FWTLLQRNQMPADKLTHEITIAAVKFTTEYAAINFINLPPTCIGELAQFYMANLILK 900
 841 FWTLLQRNQMPADKLTHEITIAAVKFTTEYAAINFINLPPTCIGELAQFYMANLILK 900

901 YCDHGOYLINTLTSITGARRPRDPSVLHWRKDVTSAAIDTQAKALEKTELPELW 960
 901 YCDHGOYLINTLTSITGARRPRDPSVLHWRKDVTSAAIDTQAKALEKTELPELW 960

961 TTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNRVFAQGNWSGLNGKVCPLTFD 1020
 961 TTAFTSTHLVRAAMNQRPMVVLGISISKYHGAAGNRVFAQGNWSGLNGKVCPLTFD 1020

1021 RTRRTIACPRGGFCPTVGPSSGNRETTLSDOVRGLIVSGAMVOLAIYATVAVRAGAR 1080
 1021 RTRRTIACPRGGFCPTVGPSSGNRETTLSDOVRGLIVSGAMVOLAIYATVAVRAGAR 1080

1081 AQHMAFDWLSLTDDEFLARDLELHDOIIQTLETPTWVEGALEAVKILDEKTTAGDGET 1140
 1081 AQHMAFDWLSLTDDEFLARDLELHDOIIQTLETPTWVEGALEAVKILDEKTTAGDGET 1140

1141 PTNLAFPDSCPSHDTTSNVLANISGNSISGSTVPGIKRPPEDDELFDLSGIFIKHGNIT 1200
 1141 PTNLAFPDSCPSHDTTSNVLANISGNSISGSTVPGIKRPPEDDELFDLSGIFIKHGNIT 1200

1201 MEM 1203
 1201 MEM 1203

RESULT 2
 AAE25598
 ID AAE25598 standard; Protein; 1452 AA.
 XX AAE25598;
 AC
 XX
 XX
 DT 04-NOV-2002 (first entry)
 XX
 DE Herpesvirus ICP8-GFP fusion protein.
 XX
 KW ICP8 protein; herpesvirus; green fluorescent protein; GFP; infection;
 XX antiviral agent; fusion protein.
 XX
 OS Chimeric - Herpesvirus.
 OS Chimeric - Unidentified.
 XX
 US6399354-B1.
 FN
 XX
 PD 04-JUN-2002.
 XX
 XX 31-JUL-1998; 98US-0127227.
 XX
 XX 31-JUL-1998; 98US-0127227.
 PR (HARD) HARVARD COLLEGE.
 XX
 XX
 PI Xnipe DM, Taylor TJ, McNamee EE;

PT viral protein derived from virus and detectable protein -
 XX Disclosure, Fig 7; 30pp; English.

CC The invention relates to a fusion protein comprising a viral protein with
 CC at least one function and a detectable protein, fused to maintain the
 CC open reading frame. The fusion protein is useful for determining whether
 CC a cell is a virus-resistant cell or a virus-susceptible cell by
 CC contacting a virus that expresses the fusion protein with the cell to be
 CC tested, under conditions sufficient for the virus to infect the cell, and
 CC detecting the presence or absence of the fusion protein, where the
 CC absence of the fusion protein identifies a virus-resistant cell and the
 CC presence of the fusion protein identifies a virus-susceptible cell. The
 CC fusion protein is useful for identifying an anti-viral agent or an agent
 CC that blocks the expression of the fusion protein by contacting a virus
 CC that expresses the protein, a host cell and the agent to be tested, in
 CC conditions sufficient to allow for the virus to infect the cell, and
 CC detecting the amount of the virus present, where a decrease in the amount
 CC of virus present identifies the agent. This sequence represents an
 CC ICP8-green fluorescent protein (ICP8-GFP) fusion protein of the
 CC invention.

XX Sequence 1452 AA;

Query Match 48.5%; Score 3054; DB 24; Length 1452;
 Best Local Similarity 49.8%; Pred. No. 2.5e-288;
 Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;

QY 1 MENTQKVT--VPTGPGVY--ACRVEDLDLEISFLAARSTDSNLLPLMRNLTVS 55
 DB 1 METKPTATIKVPPGPGVYACRSEGIEL--LALLSARGSDADVAAPLVWGLTVS 58
 QY 56 KTFSSLAIVSGARTTLAGAGITLKLTHSHFVPSVFVHGKHVLPSSAAPMLTRACNA 115
 DB 59 SGFEANVAVVGGRTTGLGTAVSLKTPSHYSSVVVFGGRHLPSTQAPNLTCLER 118
 QY 116 AREFGRSCQGPVGVAVETTGAEICTRLGLSPENTILYLVTALFKAVFWCNVFLHY 175
 DB 119 ARRHFGSDVTPRPGDUKHETTCALCERLGLDRLALLYLVTGEGKAVCINNFTLH 178
 QY 176 GGLDIVHINHDVIRIPFLPVQLFMPDVRNLVDPFNTHRSIGEGVYPTFYNTGLCH 235
 DB 179 GGSQKVTIGAEVHRIPVYQLFMPDFSEVIAEPENANHRSIGENTYPLFPFNPINR 239
 QY 236 LIHDCVTAPMAVALRVNTAVAGAAHAFDENHGVLPDPDITVYFQSSSOTTTAR 295
 DB 239 LLEAVVGPAAVALRCRNVDAVARAAHAFDENHGVLPADITVTAFAEQG--KTPR 296
 QY 296 GARNDVNSTSKPSGSGFERRLASIMADTALHAEVIENTGIYETPTDIKEMPWFIGM 355
 DB 297 GGR-----DGGKGPAGGEORLASVWAGDALALESIIVSMVFDPPDTSAPLCEQ 351
 QY 356 EGTLPRLALGSTAVAGVIGAMVSPNSALYLTVEDESGMTEAKDGGFGPSFNRFYQF 415
 DB 352 DTAARANAVGAYLARAAGLVGAMVSTNSALHTEVDAGPADPKDHSK--SFYRFLV 410
 QY 416 AGPFLAANPOTDRGHVL-----SSQSGTSNTEPSVDYLYALICGAPAPLALLPYL 468
 DB 411 PGTHVAANPQVDEGHVVPFGEFRTPAVLVGGTQ--EFAGHLAMLCGFSAPLALXLYL 469
 QY 469 ERCDAGFTGGHG--DALKYVTGTFDSEI PCSCEKHTRPVCAHTTVHRLQRMRFQAT 527
 DB 470 ERCDGGVIVGRQEMVFRVADSNQTDVPCNLCTEDTRHACVHTTLMRLARHPKFAAA 529
 QY 528 ROPIGVGTWNSOYSCDPLGNVAPYLIRKPGDQTEAAKATWQDTYRATLERLFIDLEQ 587
 DB 530 RGAIGVGTWNSOYSCDVLGNVAPYAFSAALKR--ADGSETARTIMOETYRAATVRVAELET 588
 QY 588 ERLDRGAPCSSLSSVVDHPTFRILDLTARIEQTTQFMKVLVETRDYKIREGLS 647
 DB 589 LQVVDQAVTPMGRLETTITNEALHTVNNVQVVDREVEQLMRNLVEGRNFKPRDLG 648
 QY 648 EATHSMALTDFDYSAGAFCPITNPLVLRTHLAVVQDIALSQCHVCVYGOQVEGRNFRNQF 707

DB 649 EAHAVSUTLDPYACGCPCLQLLQGRSNLAVYQDIALSQCHGVFAGOSVEGRNFRNQF 708
 QY 708 PVLRRFRVLDLFGNGGIFSTRSITVTLSEG--PVSAPNPTIGODAPAGRTFDGDLARVSVEI 766
 DB 709 PVLRRVDMFNNGGELSAKITLVALSEGAALCASLTAQTPAPRESSFEGDVARVTILGFP 768
 QY 767 RDIRVKNRVFSGNCTNLSEARARLVGLASAYQRQKRVDMHGLGFLKQPHGLLFP 826
 DB 769 KELRVKSVLFAGASANAEEAKARVASLQSAQYQKPKGVVDILGLPLGFLLKQPHAAIFP 828
 QY 827 RCMPPNSKSNPQWFTLLOKNOWPADKLTHEEITTLAAVKRFTTEEYAAINFINLPFCI 886
 DB 829 NGKPPGSGNQPPQWFTALQKQPARILSREDIETAFIKGFLDYGAINFINLAPNV 888
 QY 887 BELAOFYMANILKYCDHSOYLINTLSITGARRPRDPSSVLHWIRKDVTSADITQA 946
 DB 889 SELAMYYMANQILRYCDHSTVFINTLITAGSRPPSVQAAAAW---SAQGGAGLEAGA 945
 QY 947 KALLEKTENLDELATTAFTSTHLVRAANQRPVWVIGISIKYHGAAGNRVFOAGNWSG 1006
 DB 946 RALMDAVDAHPGAWTSMFASCNLLRPFYMAARPMVVLGLSISKYGVAGNDRVPQAGNWS 1005
 QY 1007 LMGKNVCPLETFDTRRRFIACPRGGFICPVTGPSSGNETTILSDOVGIIVSGGAMVQ 1066
 DB 1006 LMGKNVCPLETFDTRRRFIACPRGGFICPVTGPSSGNETTILSDOVGIIVSGGAMVQ 1065
 QY 1067 LAIVATVVRVAGAPQAMAFDDMLSLTDDDFLARDLELHDQIQTLETPTWVEGAL--- 1123
 DB 1066 SSVFVATVKSIGPPTQQLQIEDMLALEDEYLSSENMELTARALERGNGEWSDDALEVA 1125
 QY 1124 -EAVKILDEKTTAGDGTPTNLAFNED--SCEPSHDTTSNVLNISGNSISGTVPGIKP 1180
 DB 1126 HEAEALVSQGNAGE-----VFNFGDFCB-----DMATPFGGAGPAPAFGRKPA 1173
 QY 1181 PEDDELFDLGSIPKIGNITWEM 1203
 DB 1174 FHGDDPFG-EGPPDKGDLTLDL 1195
 RESULT 4
 ID AAW72105 standard; Protein; 1196 AA.
 AC AAW72105;
 XX 18-DEC-1998 (first entry)
 DT HSV-2 strain SB5 Contig ID 12 ORF#2 protein.
 DE HSV-2 strain SB5; immunological response induction; therapy;
 KW antiviral identification; viral protein inhibitor.
 OS Herpes simplex virus type 2.
 XX W09820016-A1.
 XX 14-MAY-1998.
 XX 31-OCT-1997; 97MO-US20016.
 XX 09-JUN-1997; 97US-0049018.
 XX 04-NOV-1996; 96US-0030279.
 XX (SMIK) SMITHKLINE BEECHAM CORP.
 XX Chan JY, Dabrowski-Amarel CE, Delvecchio AM, Dillon SB;
 XX Esser KM, Leary JJ;
 XX WPT; 1998-286847/25.
 XX DR N-PSDB; AAV62156.
 XX Herpes simplex virus type-2 sequences - useful in, e.g. prevention

PT and treatment of infection or inducing immunological response in
XX mammal

PS Claim 10; Page 83-84; 748pp; English.

XX This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC sequence of the invention. This sequence was isolated from a HSV-2 strain
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 12.
CC The proteins can be used for the treatment or prevention of disease, to
CC induce an immunological response in a mammal or to identify inhibitors,
CC activators or novel antivirals. Antagonists of the proteins can be used
CC to inhibit a viral polypeptide. The DNA sequence or a vector containing
CC it can also be used to induce an immunological response in a mammal.
XX Sequence 1196 AA;

Query Match 47.9%; Score 3016.5; DB 19; Length 1196;
Best Local Similarity 49.1%; Pred. No. 8.4e-285;
Matches 596; Conservative 211; Mismatches 367; Indels 41; Gaps 16;

4 TQKTVPTGPGYVY--ACRVEDLDLEISLAARSTDSIALPLMNLTKVETFTSS 61
7 TTTTVKVPPEYVYGRACPAEGL--LSLSARSAGDADVAVAPLVGLTVSGFEAN 64
62 LAVSGARTTLAGAGITLTKTSHFVPSVVFHGGKHVLPSSAAPNLTRACNAARBERF 121
65 VAAVVGSRITLGGTAVSLKMPHSYVSVYFHGHRHLAPSTQAPNLTRCEARRHFG 124
122 FSRCCGPPVDGAVETGABICRRLGLEBENTILYLVVLTALPKAVFMCVFLHYGGLDIV 181
135 FSDYAPRPCDLKHETTGALCERLGLDPRALLYLIVTEGFREAVCISNTFLHGLGMDKV 184
182 HINHGDIVIRIPFPVQLEMPDVRNRPVPPFNTHRSIGEGVYPTPTVNTGLCHLHDCV 241
185 TIGDAEVHRIPVYPLQMPDPSRVIAAPFNCNHSIGENFYLPPFNRLPLALLFEAV 244
242 IAPMAVALVRNVTVAVAGAAHLAPDENHEGAVLPDITYTFYSSSGSTTTARGARRND 301
245 VGPAAVALRARNVDVAAAAHLAPDENHEGAAFPADITTAFAESQG--KPGSGAR-- 299
302 VNSTSKPSPSGFERELASIMAAITALKAEVIFNTGIYEEPTDIKENPMFIGNGLR 361
300 --DAGNKPGAGFQRLASVAGDAALESIVNAVDEPPDITWPLLEGQETPAAR 357
362 LNALGSYTVARVAGVIGAMVFNAGLYTEVEDSGMTEAKDGGPSPFNRFYQAPGPHLA 421
358 AGAVGAYLARAAGLVGAMVFNAGLYTEVEDSGMTEAKDGGPSPFNRFYQAPGPHLA 416
422 ANPQTRDRGHV-----SSQSTGSSNTEFSVDYLLALICFGAPLLARLLFYLERCDAG 474
417 ANPQDRGHVVPVGVGPRPTAPLVGGTQ--EPAGEHLLAMLCGFSALLAKMLFYLERCDGG 475
475 AFTGGHG--DALKYVTGTDFSDIPCSLCEKTRPCANTVHRLRQMRPRFGQATQPIGV 533
476 VIVGQENDVRYVADSGQTVPCNLCTFETRHACATLMLRLRHPKFAASARGAIGV 535
534 FGTWNSQVDCDPIGNVAPYLLIKPKGQDETAATQDTRATLERLFDLQERLLDR 593
536 FGTWNSAYSDCDVLGNVAAAFSAKX--ADGSENTRTIMQETVRAATERVMALEALQVQD 594
594 GAPCSSEGLSVIVDHPFRILDTLRAEIQTTQFMKVLVETRDYKIREGLSEATHSM 653
595 AVPTALGRLEITLITREALHTVNNIKQLVDREVEQLMRNLIEGRNPKFRDGLAEANHAM 654
654 ALTPDPXSGAFCPITNLFVRLKTHLAVVQDLALSCHVYFGQVQVQVGNPNRPNQFVLRR 713
655 SLSDPVTGCPPLQLLARRSNLAVVQDLALSCHVYFGQVQVQVGNPNRPNQFVLRR 714
714 FVDLFGNGFTSTRITVTLSEG--PVSNPNTLGDAPAGTFCGDLARVSVEVIRDIRVK 772
715 VMDLFFNGFLSAKTLTVALSEGAICAPSITAGTAPAESSEFGDVARVTLGPFKELRVK 774
773 NRVVFSNGCTNLSEARARLVGLASAYQORKEKRVMDLHGALGFLLLKQFHGLLPPRGWPN 932

Db 775 SRVLFAGASANAASEAAKARVASLSQSAVQKPRVDILLGLPLGLLKQFHAVIFPKPG 834
QY 833 SKSPNCPWFLLQRNCPADKLTHEITITIAAVKPTTEYAAINFINLPPTCIGELAQF 892
Db 835 SNQPNQWFTALQRNQLPARLLSREDIETIAFKFPLSDYGAINFINLAPNNVSLAMY 894
QY 893 YMANILKYCDHSQYLINTLTSITGARRPRDPSVLHWIRKDVTSAAADIETCAKALLEK 952
Db 895 YMANQLLYCDHSTYINTLTVIAGSRPPGVQAAAAPQ---CGAGLEAGARALMDS 951
QY 953 TENLPELWITAFSTHLVRAAMQRMVVLGISISKYHGAAGNNRVFOAGNWSGLNGKN 1012
Db 952 LDAPGAWTSMFASCNLLRPVAAAREPMVVLGLSISKYGVAGNDRVFOAGNWSLIGKN 1011
QY 1013 VCPLEFTRTRRRIIACPRGGFCPTVGPSSGNRETTLSDOVRGIIVSGAMVQALAIYAT 1072
Db 1012 ACPLLFTRTRKVLACPRAGFVCAASSLGGGAHSHLCEQURGIAGGAAVASSVFA 1071
QY 1073 VYRAGARQAHMAFDWLSTLDDFLARDLELHDQIIQTLTPMTVEGAL---EAVKI 1128
Db 1072 TVKSLGPTQQLQIEDWLALLEDEVLSSEEMFTTRALERGHGEWSTDAALVAAHEAEAL 1131
QY 1129 LDEKTTAGDGETPTNLAFNFDSCESHDTTSNVLAISGNSISGTVPGIKRPPDELEFD 1188
Db 1132 VSQGAAGE-----VFNFGDGEDDDHAASFGGLAAA--AGAAAGVARKRAFHDGDPFG 1182
QY 1189 LSGIPIKHGNTIEM 1203
Db 1183 -EGFPEKK-DLTLD 1195

RESULT 5

AAW72197
ID AAW72197 standard; Protein; 623 AA.

XX AAW72197;

XX 13-JAN-1999 (first entry)

DE HSV-2 strain SB5 Contig ID 15 ORF#32 protein.

XX HSV-2 strain SB5; immunological response induction; therapy;
XX antiviral identification; viral protein inhibitor.

XX Herpes simplex virus type 2.

XX WO9820016-A1.

XX 14-MAY-1998.

XX 31-OCT-1997; 97WO-US20016.

XX 09-JUN-1997; 97US-0049018.

XX 04-NOV-1996; 96US-0030279.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Chan JY, Dabrowski-Amara CE, Delvecchio AM, Dillon SB;

XX Esser KV, Leary JJ;

XX WPI; 1998-286847/25.

XX N-PSDB; AAV62176.

XX Herpes simplex virus type-2 sequences - useful in, e.g. prevention
XX and treatment of infection or inducing immunological response in
XX mammal

XX Claim 10; Page 123; 748pp; English.

CC This sequence represents a Herpes simplex virus type-2 (HSV-2) protein
CC sequence of the invention. This sequence was isolated from a HSV-2 strain
CC SB5 (deposited as ATCC VR-2546) DNA fragment designated Contig ID 15.

Db 478 LVSQGAAGE-----VFNGDFGDEDDHAASFGGLAA--AGAGVARKAFHGDOPF 528

2y 1188 DLSGIPKIGNITMWM 1203

Db 529 G-EGPPEKK-DITLDM 542

RESULT 7

AB53126

AB53126 standard; Protein; 1132 AA.

AB53126;

28-FEB-2001 (first entry)

Macaca mulatta rhadinovirus 17577 RRV ORF6 protein SEQ ID NO:9.

Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus; genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6; IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine; cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma; lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy; splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia.

Macaca mulatta rhadinovirus 17577.

W0200028040-A2.

18-MAY-2000.

05-NOV-1999; 99NO-US26260.

06-NOV-1998; 98US-0107507.

20-NOV-1998; 98US-0109409.

(UYOR-) UNIV OREGON HEALTH SCI.

Wong SW, Axthelm MK, Searles RP;

WPI; 2000-376552/32.

New rhesus rhadino virus for producing non-human primate model useful for testing potential treatments and efficacy of the candidate vaccine for conditions associated with RRV infection

Claim 5; Page 123; 141pp; English.

The present invention describes a novel rhesus macaque rhadinovirus called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the RRV genome sequence, and AAB53204 represent the proteins encoded by the genome sequence. The present invention also specifically claims the individual open reading frame (ORF) nucleotide sequences from the genome which encode the individual proteins, but these sequences are not given. A non-human animal infected with RRV can be used for testing the efficacy of drug in the treatment of condition associated with infection with RRV such as Kaposi's sarcoma, lymphoproliferative disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly, hypergammaglobulinemia or autoimmune haemolytic anaemia, by administering the drug to an immuno-compromised non-human primate preferably Rhesus macaque monkey obtained by as a result of infection by Simian Immunodeficiency Virus (SIV). RRV is useful for producing non-human primate model for testing potential treatments for conditions associated with RRV infection. It is also useful for testing the efficacy of the candidate vaccine against RRV infection or conditions associated with its infection by administering the vaccine to the subject capable of infection with RRV, inoculating the subject with RRV and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205 to AAB53213 represent sequence used in the exemplification of the present invention.

Sequence 1132 AA;

Query Match 9.9%; Score 620; DB 21; Length 1132;

Best Local Similarity 22.6%; Pred. No. 2.7e-50;

Matches 290; Conservative 212; Mismatches 506; Indels 274; Gaps 50;

Qy 1 MENTOKTVPTPLGYVYACRVEDLDLEISFLAARSTDSLDLPLMRNLVTEKTFYS 60

Db 11 LEDNQS-RAPICAGCYVAYSQDFFFAEASTLGNRPSGVSFLPILYGLTVEHEFL 69

Qy 61 SL-AVVSAGARTTGLAGAGITLKLTTSHFYPSVVFHGGKVLFPSSAANLTRACNAAR 119

Db 70 TVKAAYKKVDTTTTLA-----VKYTC--FHREIVTFHNASLFRPVPDGTGLNELCEARAL 122

Qy 120 FGSRC--QOPP--VDGAVETTGAEICTRLGLEPENTILYLVVTLFKE-----AV 166

Db 123 FGYQTFIEPGPHSIWNPLE-----CPQLP-DKDEMFLGVVVTEGFKERLWRGCLVPAV 175

Qy 167 FMCNVFLHYGGLDIVHINEGDVIRIPLFPVQLFMPDVNRLVDPDPNTHERSIGEGVYPT 226

Db 176 FQYQQ-----VQIARQAFKFLYDEDFAPRHNM-----P 207

Qy 227 PFYNTGLCHLIHDCVIAPMAVALRVNVTAVARGAAHLAFDENHEGAVLPDITVYFQS 286

Db 208 RFYHKDVSAYLVDSLFTSIAQALRLKDVAVIHATEKQPMODHYKIAXI---VQAKQF-- 262

Qy 287 SSSGTTTARGARENDVNSTSKSPSGGFERRLASIMAADTALHAEVIFNTGIYEETPTDI 346

Db 263 -----STLPKTTDGGSHMTVDVSVVLAELALSYGCMF-----LSCPQDA 300

Qy 347 KE-----WPMFIGMEGLTFLRLNALGSYTAHVIGVIGAMVFPNSALYTEVEDSGMTEA 400

Db 301 CELLYDVSWPIFDGCDSPPEARVNALERSAEQAVHAGQLPAANSVLVLTVKQCAPRQ 360

Qy 401 KGGPQPSFNRYQFAGPHLAANPOTDRGH-----VLSQSTGSSNTEFSVDYIALICG 455

Db 361 K--GDVNVNSFFLOHGLFLNEATIKENGSAFKGVPSNALDGSFTPY---HLAYAAS 415

Qy 456 FGAPLIARLLFYLERCDAGFTGGHGDALKVVTGFDSEI PCSLCEKTRPVCAHTTVHR 515

Db 416 FSPHLLAKLYNQFLOHKKSTNQAFNMVHVYVTAANSEWCTLCHGNTPATCLNTLYR 475

Qy 516 LRQRPFRQATQ--PIGVFGTMNSQYDCDPLGNAPYLILRKEGDOT-FAAKATMODT 573

Db 476 LKDRPFAVTPORRDPYVVTGTAGT-FNDLEILGNFASFRDREEDGNPADEHPKYVYQ 534

Qy 574 YRATLERFLDLQERLLDRGAPCSSEGLS-----SVIVDHTPFRILTLRARIQT 626

Db 535 CQVTEKL-----SAIGITEDHNVNLTNIQSFLRVFKGDISVDGE 578

Qy 627 TTQFMKVLVETRDYKIREGLSEATHSMALTDPYSGAFCPITNPLVKRTHLAVVQDLALS 686

Db 579 VMKFNVMK--NNFNREHVKSVHILQFCNVYVQAPCAVFLNLYKSLWIIQDCLP 637

Qy 687 QCHCVFYQ-----QVEGNFRN-----QFQVLRFRFVDLF 718

Db 638 --YCMIVQDDNPAMGILPSEWLKMHFOTLWNTNFAACLDRLGVLTCGLKIVHRDMFCDF 695

Qy 719 -----NGQFISTRTITVLSGPPVSNPTLQDAPAGRTFDGLARVSVVEIRDVKN 773

Db 696 DTDAGSNGMLAPFKQVVR-----IARMMVVPKSIKKN 729

Qy 774 RVVFGNCTNLSAARAPLV--GLASAYQKQKRVDMHAGLGLLKQPHGLLFPKGM 830

Db 730 RIIFEN--TAGSEAVQSGFVKPTGTRDTY-----VAGPYMKFLNSLHRALE----- 774

Qy 831 PMSKSPNPOWMTLLQRONPADK-LTHEETITIA--VKFTSEYAININLPPTGICE 888

Db 775 PDTKTAALYHKKISQNKTPVLKDVDPDELAELVSVKVNLSLAFTETNVDVPSLMS 834

Qy 889 LAQFYMANLILKYCDHSQYLINTITSITGARPRDPSSVLHMRKDVTSAAADIETOKA 948

Db 835 YARIKNGAILRACQIQFYATTL-HCLTPVQIDAEYPH-----VLGSAIAIPVAY 886

Qy 949 LLE---KTENLPWLWTFTASTHLYVRAMNQRPNVLGISIKTHGAGNVRVFOAGN-- 1003

Db 889 LAEIRGRT-----ALTQTTARQPVAAGRLRPVITVPMVKNKTVGNVNNVHFCGNLG 943
 QY 1004 -WSGLNGKNVCPLFT-FDRT-----RRFIACPRGGFTCPVTGPSSGNRETTLSQDV 1054
 Db 944 VPAGGVDRNLWPSSPKKTVGSAMLRKRHMVMTTP-----IIDRLIKRAAG--QTISTFEA 998
 QY 1055 RGIIVSGGAMVQ-----LAIYATVVRVAVGARAQHAEPDDWLSLTDDEFLARDLEELH 1106
 Db 999 ESKVRSVQALLKEDKNPNLLKSVILELRHKGKQD-----LSEEDVQYLYGDYCMLT 1052
 QY 1107 DQIQITL-----TPWTVGEALRAVKILDEKTTAGDGETPTNIAFNFDSCERSHD---T 1157
 Db 1053 DEVLFTLDNIAQSGVFWTIE---DAGALIEDRQADDLQFVDSDDIATASCQPEELPT 1109
 QY 1158 TSNVLNIGSNGISGTVPLKX 1179
 Db 1110 PS-----AGALLACKX 1121

RESULT 8
 AAY26935
 ID AAY26935 standard; peptide; 35 AA.

AC AAY26935;
 XX

DT 21-DEC-1999 (first entry)
 XX

DE Varicella zoster virus type 29 recombinase D35E motif.
 XX

KW Recombinogenic motif; retrovirus; integrase; invertebrate; transposase;
 KW consensus; vaccine; gene therapy; allergy; autoimmune disease; cancer;
 KW cardiovascular disease; graft rejection; hematopoietic disorder;
 KW immunosuppressive disorder; immunoproliferative disease; jaundice;
 KW immunodeficiency disease; infectious disease; inflammatory disease;
 KW septic shock; metabolic defect.

XX Varicella zoster virus.
 XX

OS US5959074-A.
 XX

PN 28-SEP-1999.
 XX

PD 28-FEB-1997; 97US-0807332.
 XX

PF 01-MAR-1996; 96US-0012616.
 XX

PR 02-AUG-1996; 96US-0023064.
 XX

XX (NAJE-) NAT JEWISH MEDICAL & RES CENT.
 XX

PA Gelfand EW, Dreyfus DH;
 XX

PI WPI; 1999-561074/47.
 XX

DR Peptides derived from a Herpes virus recombinogenic motif useful for the
 XX production of vaccines and in gene therapy -
 XX

PT Claim 1; Column 37-38; 29pp; English.
 XX

PS The invention relates to novel recombinogenic motifs found in
 XX (retro)viral integrases which have similarity to invertebrate
 XX transposase molecules. The peptides, or their homologues, have the
 XX following identifying characteristics: (a) an initial and terminal amino
 XX acid comprising Asp or Glu, separated by 34-39 amino acids, especially
 XX 35 amino acids (the D35E motif); (b) a high probability (p < 0.05) of
 XX alignment with a consensus sequence (AAY26941), as determined by:
 XX (1) aligning the initial and terminal residues of the homologue sequence
 XX with those of the consensus; (2) aligning the intervening residues of
 XX the homologue with those of the consensus by maintaining the spacing of
 XX the consensus while, if necessary, altering (by the insertion of spaces
 XX or the deletion of residues) the spacing of the homologue; and
 XX (3) generating a distribution of 1500 random sequences of amino acids
 XX between the initial and terminal residues and identifying the percentage
 XX of aligned residues. The recombinogenic motifs may be useful in the

CC production of vaccines and in gene therapy to treat allergies,
 CC autoimmune diseases, cancers, cardiovascular diseases, graft rejection,
 CC hematopoietic disorders, immunosuppressive disorders,
 CC immunoproliferative diseases, immunodeficiency diseases, infectious
 CC diseases, inflammatory diseases, jaundice, septic shock and other
 CC immunological, genetic or metabolic defects.

XX Sequence 35 AA;
 SQ

Query Match 2.9%; Score 183; DB 20; Length 35;
 Best Local Similarity 100.0%; Pred No. 3.5e-10;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 545 DPLGNYPYLIRKPGDQTEAAKATMDTYRATLE 579
 Db 1 DPLGNYPYLIRKPGDQTEAAKATMDTYRATLE 35

RESULT 9
 AAB73931
 ID AAB73931 standard; peptide; 35 AA.

XX AAB73931;
 AC

XX 29-MAY-2001 (first entry)
 DT

XX Varicella zoster VZV29 protein recombinogenic sequence.
 DE

KW Varicella zoster virus; VZV29; herpesvirus; D35E; anti-viral;
 KW immunosuppressant; immunostimulant; cytostatic; anti-allergic; cardiant;
 KW anti-inflammation; gene therapy; infection; herpes virus replication;
 KW V(D)J recombination; retroviral integrase;
 KW immunoglobulin class switching; allergy; autoimmune disease; cancer;
 KW cardiovascular disease; graft rejection; haematopoietic disorder;
 KW immunological disorder.

XX Varicella-zoster virus.
 OS

XX US6187584-B1.
 PN

XX 13-FEB-2001.
 PD

XX 23-JUN-1999; 99US-0338876.
 PF

XX 01-MAR-1996; 96US-0012616.
 PR

XX 02-AUG-1996; 96US-0023064.
 PR

XX 28-FEB-1997; 97US-0807332.
 XX

XX (NAJE-) NAT JEWISH MEDICAL & RES CENT.
 PA

XX Dreyfus DH, Gelfand EW;
 PI

XX WPI; 2001-256131/26.
 DR

XX New nucleic acid molecules encoding peptides with transposase
 PT activities, useful as immunosuppressants, anti-viral agents, or vectors
 PT for gene therapy of diseases, e.g. allergies, autoimmune diseases,
 PT cancers or graft rejection -
 PT

XX Claim 1; Column 25; 31pp; English.
 PS

XX The present sequence is provided in a specification relating to a new
 CC isolated nucleic acid molecule comprising a nucleic acid sequence
 CC encoding a peptide consisting of one of ten amino acid sequences
 CC defined in the specification or homologues of these sequences.
 CC The nucleic acid molecules are useful for producing peptides.
 CC important in the regulation and function of herpes virus replication,
 CC V(D)J recombination, retroviral integrase function or immunoglobulin
 CC class switching. The isolated nucleic acid molecules and the peptides
 CC they encode are useful as immunosuppressants, anti-viral agents or
 CC vectors for gene therapy. These are particularly useful in protecting
 CC humans from a variety of diseases, e.g. allergies, autoimmune diseases,
 CC cancers, cardiovascular diseases, graft rejection, haematopoietic

Db	2631	-PVISDLAVTLRQGLTLENAIPATPDSGYWEQTMSPATTPSDMEAVLCKEPAVLG---	2688
Qy	466	FYLERCDAGATGGHG-----DALKYVTGTFDSEIPCSLCEKHTRPVCAHT	511
Db	2687	VEVSATDSFDLGGHSLMATKLAARISRLDVPVSIKIDFHSVPLNLARK-----	2737
Qy	512	TVHRLRQRPRFQCAATRQPIGVFGTWNQSYDCDPLGNVAPVLIL-----	561
Db	2738	-----IRLTQAKGHEATN--GV-----QIANDAPQLISVEDPEIFVQREIAP	2778
Qy	562	QTEAAKATWQDYVRAT-LERLP-----IDLQERLLDRGAPCSS--EGL	602
Db	2779	QLQCSPEITILDVYPATQMORVELLNEVTKPRSPTEPHIDFPD-----	2832
Qy	603	SSVIVDHPTRFRILDLTRARIEQTITTFMKVLVEVDYKIRGLSEATHS-----	652
Db	2833	ASLAKHFDIFRIVLEARGELQVVLKHVDVPIEM--LOTEENINSATRSFLDVAEKPI	2890
Qy	653	-----MALTFDPYSGAFCPITNPLVXRTHLAVVQDLALS-----	699
Db	2891	RLGQPLIRIAILEKPGS-----TLRVLRLSHALYDGLSLEHILSLHLFFGGSLPPP	2944
Qy	700	RIFRNFQFQVL--RRRFVULFNGGFISTRASITVTLSEGPVSAPNPTLGODAPAGRTFDGD	757
Db	2945	PKFAGYMQHVASSRREGYDFWR-SVLRDSSMTVIKGNNTTTPPPPPQQSTPSGAHHASK	3003
Qy	758	LARVSVEVIRDRVRKNRVVFGNCT-----NLSEAAARARLV-----	800
Db	3004	VVTIQTQNTDSRIETATIFTTACALMLAKEDNSDVVFGRTVSGROGLPLAHQ	3057
RESULT 13			
ID AAY00206			
AAAY00206 standard; Protein; 1074 AA.			
AC	AAAY00206;		
XX	XX		
XX	DT	20-APR-1999 (first entry)	
XX	XX		
DE	XX	Enterococcus faecalis protein EF102.	
XX	XX		
KW	XX	Enterococcus faecalis; infection; vaccine; immune response; diagnosis;	
KW	XX	detection; attenuation; antigenic.	
XX	XX		
OS	XX	Enterococcus faecalis.	
XX	XX		
PN	XX	W09850554-A2.	
PD	XX	12-NOV-1998.	
PF	XX	04-MAY-1998; 98WO-US08959.	
PR	XX	14-NOV-1997; 97US-0066009.	
PR	XX	06-MAY-1997; 97US-0044031.	
PR	XX	16-MAY-1997; 97US-0046655.	
XX	XX	(HUMA-) HUMAN GENOME SCI INC.	
XX	XX		
PI	XX	Bailey C, Choi GH, Hromockyj A, Kunsch CA;	
PI	XX		
DR	XX	WPI: 1999-070095/06.	
DR	XX	N-PSDB; AAX20196.	
XX	XX		
PPT	XX	New isolated Enterococcus faecalis polynucleotides - used to develop	
PPT	XX	products for the detection of Enterococcus and for use in vaccines	
PPT	XX	for prevention or attenuation of Enterococcus infection	
XX	XX		
PS	XX	Claim 9; Page 197-198; 30pp; English.	
XX	XX		
CC	XX	The present sequence represents a protein isolated from	
CC	XX	Enterococcus faecalis. The present invention describes genes, proteins	
CC	XX	and antigenic polypeptides isolated from E. faecalis. The proteins can	

CC be used in vaccines for preventing or attenuating an infection caused
 CC by a member of the Enterococcus genus in an animal. They can also be
 CC used for detecting Enterococcus antibodies in a sample. The nucleotide
 CC sequences can be used for detecting Enterococcus nucleic acids.
 CC Products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of E. faecalis protein
 CC activity.
 XX
 SQ Sequence 1074 AA;
 Query Match 2.0%; Score 123; DB 20; Length 1074;
 Best Local Similarity 18.3%; Pred. No. 0.11;
 Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;
 QY 250 RVNVTAVARGAHLAPDENHEGAVLPDILTYFQSSSSGTTTARGARNVNSTSKPS 309
 Db 67 RTTSIAYENGAKQTVFC-IEPGVSIPEVTHGY-----QRNPL 104
 QY 310 PSGPERLASIM--AATLHAIEVFNTHGYEE-----TPTDIKEWPMFI 353
 Db 105 PMSDKAKLVSVLWEKAGTDIDTNVAKMIWEVGVKLSIKRLGASVDIK----- 158
 QY 354 GMEGTLPRNLALGSYVAR-----VAGVIG-AMVFPNSALYITEVEDSGMTEAKDGGP 405
 Db 159 SIEGKINK--AIEYQKPSFHNTVTKILGQSTLIDKNELNSEFDKVVQNTA----- 211
 QY 406 GPSNRYQFAGPHLAANPQTRDGHVLSGSGTSSNTFSDVYALICGFGAPLLARLL 465
 Db 212 ---NIDYRVIGNQLVLP-----NSNSKSGTLTKKSAGTGP-VAYKK 251
 QY 466 FYLERCAGFTGCHGALKVVTTFSEIPSCIEKHTRPVCAHTVHRLORMPRFQG 525
 Db 252 AGLQTVNAGALDKNTVAIKNVETKGS-LKIKIDKESGDIVPETHL-----DFGK 304
 QY 526 A-----TROPIGVGMNSQYSCDPLGNVA-----PYLILKPKQDQTEAAKATM 570
 Db 305 ALPSKDVTTDKGI-----SILDGIPGHYKVIKESVDPDPVMDITPMAATIKAGETI 358
 QY 571 QDTRATLERLFDLEORLLDRGAPCSSEGLS-----SVIVDHPT---FRILDTLRA 621
 Db 359 SMTSKNRQGOILLKXTG-VETGTDLWNDVSLAGNTFAIRKDSPAGEIVQBITDEKG 417
 QY 622 RIEOTTQFMKVLVETRDYKIRGLSATHSMALTRDP-----YSGAPCPTNPLVKT 675
 Db 418 RAE--TPKELANALELGTYYVTE--TKSSNGFNTFKPTKVELKYNQVALVTSNVKQ 473
 QY 676 HLAUVQDLAL-----SQCHCVFYGOV-----EGNFRNQFPVLRFRFVDL 717
 Db 474 NQBITGETTLTKEDKOTGNEQKAEFGAEYTLFTAKDQAVKWEAFK-----TEL 526
 QY 718 FNGGFTSTRITVLSB-GPVSAFNPETLGO-----DAPAGRTFDGLARVVE----- 764
 Db 527 VKGTASDETIVTLALDEKQAVKHLAINEYFWOETKAPEGYTLDETRKYPVSIKKVDNE 586
 QY 765 ---VIRDIRVKNRV-----FSGNCTNLSEAA-----RARLVGLASAYQREKR 805
 Db 587 KNAVITRDVTAKEQVIRFGDFDKFAGSADGTAETGFDNLSFKVSPLEGTXITGAEDA 646
 QY 806 VDLHLGALGF-----LKQPHGLLFPFGM-----PPNSKS-----PNP 838
 Db 647 TTACNEQLGSDGYGKFPENLPYGYLLIEIEA---PEGFKITPLEIRSTFFKKNDDYAKS 703
 QY 839 QWFTWLLQRNQ-----MPADKLTHEITITIAVKRETEYAINFNLN-PTCIGE 888
 Db 704 EYVFTITEGQKQPKMVTVPFEKLTNNE-----FVSLNRLMLYDLPEKEDSITS 754
 QY 889 LAQFYVANLILKCDHSQVILNLTSLIITGARRRPPSSVLHWIRKD---VTSAAIDETQ 945
 Db 755 LATWKGNGKLANLDFTE-LVDKL-----RVNLHEIKEDWYVAQALDVEA- 799
 QY 946 AKALLEKTENLPELWTTATFTST-----HLVPAAMNQPMVVLGISIKYHGA 992
 Db 800 TKAAQEKDEKAPVIAETATLANKEKGTGWKILFKLTAEO-----VLDKSIVLFNIV 853

QY 993 AGNRVFOAGNWSGLNGKVKVCLPFTEDTRFIIACPRGGFICPVTPGSSGNRETTILSD 1052
 Db 854 YENKVAFEAGNE-----PVA-----KDAISLN 875
 QY 1053 QVRGIIVSGAMVQLAIYATVVRVARGAR-----AQHMAFDDWLSTDDDEFIARDL-- 1102
 Db 876 QAQ-----TVNCTIERHVSIOQKAHLEDGSGTFTHGDVMDMDFDVSVTHDVL 923
 QY 1103 --EELHDIQIQTLETPWTVEGALEAVKILDE-----KTTAGD-----GETPTNLAFNFD 1149
 Db 924 GSKEAFETILYALLPDGNTKEIWSGKIEHEVNOKFTKTVLAEKVDTGKYPEGTKFTF- 982
 QY 1150 SCEPSHDTTSNV-----LNISGNSISGSTVPGIKRPPEDDE 1185
 Db 983 -TEINYEKDGNVGKHNEDLKEKSQLTLTPKEVPTIPSTPKQPE 1024

RESULT 14
 AAY00188
 ID AAY00188 standard; Protein; 1074 AA.
 XX
 AC AAY00188;
 XX
 DT 20-APR-1999 (first entry)
 XX
 DE Enterococcus faecalis protein EF094.
 DE
 DE Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 KW detection; attenuation; antigenic.
 XX
 OS Enterococcus faecalis.
 XX
 PN WO9850554-A2.
 XX
 PD 12-NOV-1998.
 XX
 PF 04-MAY-1998; 98WO-US08959.
 XX
 PR 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;
 XX
 DR WPI; 1999-070095/06.
 DR N-PSDB; AAX20178.
 XX
 PT New isolated Enterococcus faecalis polynucleotides - used to develop
 PT products for the detection of Enterococcus and for use in vaccines
 PT for prevention or attenuation of Enterococcus infection
 XX
 PS Claim 9; Page 186-187; 30pp; English.
 XX
 CC The present sequence represents a protein isolated from
 CC Enterococcus faecalis. The present invention describes genes, proteins
 CC and antigenic polypeptides isolated from E. faecalis. The proteins can
 CC be used in vaccines for preventing or attenuating an infection caused
 CC by a member of the Enterococcus genus in an animal. They can also be
 CC used for detecting Enterococcus antibodies in a sample. The nucleotide
 CC sequences can be used for detecting Enterococcus nucleic acids.
 CC Products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of E. faecalis protein
 CC activity.
 XX
 SQ Sequence 1074 AA;

Query Match 2.0%; Score 123; DB 20; Length 1074;
 Best Local Similarity 18.3%; Pred. No. 0.11;
 Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

QY 250 RVNVTAVARGAAHLAFDENHEGAVLPDITYTTFQSSSGTTTARGARRNDVNSTKPS 309
Db 67 RTTSLVAEYNGAKQTVFC-IEPGVSIPTVTHGY-----QKNPL 104
QY 310 PSGGFERRLASIM--AADTALHAENVFNTGIYEE-----TPTDIKEWPMFI 353
Db 105 PMSDKAKLVSVLWEKAGTDIDTNNVAQKMLWEVNGYKLSIKRLGGASVDIK-----158
2Y 354 GMEGTLPRLNALGSYAR-----VAGVIG-AMVFPNSALYLTVEVDSGMTAKGGP 405
Db 159 SIEGKINK--AIEEYQKPSFHNVTVKTLGQSTTLIDKNELNLSFDFKVVQNTA-----211
2Y 406 GPSFNRFYQFAGPHLAANPQTRDCHVLSQSTGSSNTFSDVYALICGFGAPLLARLL 465
Db 212 ----NIDYRVIGNQLVLT-----NSNSKSGTTLTKSAGTGTP-VAYKK 251
2Y 466 FYLERCDAGFTGGHGDALKYVGTGFDSEIPCSLCEKHTRPVCAHTVHRLRQRPFRGQ 525
Db 252 AGLQTVAGALDKPNTYAIKINVTGKS-LKIKKIDSGSDIVPETFHL-----DFGK 304
2Y 526 A-----TROPIGVFTMNSQYSDCDPLGNYA-----PYLILRKPQDQTEAAKATM 570
Db 305 ALPSKDVTTDKGI-----SILDGIPHGKTVITEKSVDPDYMIDTTPMAATIKAGETI 358
2Y 571 QDYRATLERLFDLEOEBLLDRGAPCSSEGLS-----SVIVDHT-----FRRILDTLRA 621
Db 359 SMTSKNMROKQILLKKTG-VETGTLMDNDYSLAGNTFAIRKDSPAGEIVQEIITDEKG 417
2Y 622 RLEQTTQFMKVLVETRDYKIREGISEATHSMALTFDP-----YSGAFCPITFLVKRT 675
b 418 RAE--TPKELANALELGTYYVTE--TKSNGFVNTFKTKVELKYANGTVALVTSNVKGQ 473
Y 676 HLAUVQDLAL-----SOCHCVFYQQV-----EGRNFRNQFQVLRARRFVDL 717
b 474 NOBITGETTLTKEDKDTGNESQKAEFGAEBYTLFTAXDGOAVKWEAFK-----TEL 526
Y 718 FNGGFISTRSIVTLSE-GPVSAPNPTLGO-----DAPAGTFDGLARVSVE-----764
b 527 VKGTKASDETIVLADENQNAVHKLAINFVWETKAPBGTTLDETYPVSIKKVDNNE 586
Y 765 ----VIRDIRVKNRVV-----FSGNCTNLSEAA-----EARLVGLASAYQOROKR 805
b 587 KNAVITRDVTAKEQVIRGDFDFPKFAGSADGTAETGFNDLFSKVPSPLEGTYXETGADKA 646
Y 806 VDMHLGALGF-----LLKQFHLLPFRGM-----PNPSKS-----PNP 838
b 647 TTACNEQGLFGYKGFENLPYGVDLLEBIEA---PEGFQKITPLEIRSTFKENKDDYAKS 703
Y 839 QWFMTLLQRNQ-----MPADKLTBEEITIAAVKRFTSEYAAINFILP-PTCIGE 888
b 704 EYVFTITEEGQKQPKVMVTVYEXLITNE-----FVSLNRLMYDLPEKEDSLTS 754
Y 889 LAQFYMANLILKYCHSQYLINTLTSITGARRPRDPSSVLHWIRKD---VTGAADIETQ 945
b 755 LATWQGNKKLNTLDFTE-LVDKL-----RYNLHEIKEDWYVVAQAIDVEA- 799
Y 946 AKALLEKTENLPWLWTTAFTST-----HLVRAAMQRPWVLGIGISKVHGA 992
b 800 TKAAQEKDERKAPVVAETATLANKEKTGTWKILHKLTAEQ-----VLDKSIVLVNYV 853
Y 993 AGNRVTFQAGNWSGLNGKNCVPLFTFDRTRRFIACPRGGFCIPVTGPSSGNRETTLS 1052
b 854 YENKVAPEAGNE-----PVA-----KDALNN 875
Y 1053 QVRGIIVSGGMVQLAIYATVTRAVGAR-----AQNMAFDWLSTLDEFLARDL---1102
b 876 QAQ-----TVNCTIERHVSQTKAHLEDGQSTFTHGVDMDMFDVSVTHDVL 923
Y 1103 --EELHQIITQLETPWTFEGALEAVKILDE-----KITAGD-----GPTPTNLAFFND 1149
b 924 GSKEAFETILVALPLDGTNKEIKWSKGIHEVNDKSFETKVLAEKVDYTKYPGTFTF- 982
Y 1150 SCEPSHDTTSNV-----LNISGNSIGSTVPGLKRPPEDDE 1185

Db 993 -TEINYEKDNVGNKHNEDLKESQTLTPKEVFTIPSTPKQPE 1024

RESULT 15

ABP43407
ID ABP43407 standard; Protein; 1074 AA.

XX ABP43407;

XX 05-AUG-2002 (first entry)

XX E faecalis EF094 protein.

XX Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic.

XX Enterococcus faecalis.

XX US2002045737-A1.

XX 18-APR-2002.

XX 04-MAY-1998; 98US-0071035.

XX 04-MAY-1998; 98US-0071035.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Bailey C, Hromockyj A, Kunsch CA;

XX WPI; 2002-425450/45.

XX N-PSDB; ABN98163.

XX New genes and polypeptides from Enterococcus faecalis, useful as
PT vaccines for preventing, treating or attenuating an infection caused by
PT a member of the Enterococcus genus in an animal, particularly E.
PT faecalis

XX Claim 9; Page 163; 255pp; English.

XX The present invention provides the protein and coding sequences of a
CC number of polypeptides from Enterococcus faecalis. The proteins can be
CC used as vaccines for preventing or attenuating an infection caused by a
CC member of the Enterococcus genus in an animal, particularly E. faecalis.
CC The polynucleotide is also useful for preventing or treating E. faecalis
CC infection. The present sequence is a protein of the invention.

XX Sequence 1074 AA;

Query Match 2.0%; Score 123; DB 23; Length 1074;

Best Local Similarity 18.3%; Pred. No. 0.11;

Matches 206; Conservative 150; Mismatches 415; Indels 352; Gaps 52;

QY 250 RVNVTAVARGAAHLAFDENHEGAVLPDITYTTFQSSSGTTTARGARRNDVNSTKPS 309

Db 67 RTTSLVAEYNGAKQTVFC-IEPGVSIPTVTHGY-----QKNPL 104

QY 310 PSGGFERRLASIM--AADTALHAENVFNTGIYEE-----TPTDIKEWPMFI 353

Db 105 PMSDKAKLVSVLWEKAGTDIDTNNVAQKMLWEVNGYKLSIKRLGGASVDIK-----158

QY 354 GMEGTLPRLNALGSYAR-----VAGVIG-AMVFPNSALYLTVEVDSGMTAKGGP 405

Db 159 SIEGKINK--AIEEYQKPSFHNVTVKTLGQSTTLIDKNELNLSFDFKVVQNTA-----211

QY 406 GPSFNRFYQFAGPHLAANPQTRDCHVLSQSTGSSNTFSDVYALICGFGAPLLARLL 465

Db 212 ----NIDYRVIGNQLVLT-----NSNSKSGTTLTKSAGTGTP-VAYKK 251

QY 466 FYLERCDAGFTGGHGDALKYVGTGFDSEIPCSLCEKHTRPVCAHTVHRLRQRPFRGQ 525

Db 252 AGLQTVAGALDKPNTYAIKINVTGKS-LKIKKIDSGSDIVPETFHL-----DFGK 304

QY 526 A-----TROPIGVGTMSQVSDCDPLGNYA-----PYLILRKGQDQTEAKATM 570
Db 305 ALPSKDVTTDKGI-----SILDGPHGTQVITTEKSVDPDYMIDTTPMAATIKAGETI 358
QY 571 QDYATRLERLFDLEQERLLDRGAPCSSEGUS-----SVIVDHT---FRRILDTLRA 621
Db 359 SMTSKNMROKQILLEKTG-VETGTDLMNDNYSLAGNTFAIRKDSPAGEIVQEIITDEKG 417
QY 622 RIEQTTQPMKVLVETRDYKIREGSEATHSWALTFDP-----YSGAFCPITNPLVKRT 675
Db 418 RAE--TPRELALALELGYVTE--TKSNGFVNTFKTKVELKYANQTVALTVSNVKGQ 473
QY 676 HIAVVDLAL-----SOCHCVFYGOV-----EGNPRNQFQVLRERRFVDL 717
Db 474 NQEIITGETTLTKEDKDTGNESQKAEFKGAETLFTAKDGOAVKWEAFK-----TEL 526
QY 718 FNGGFISTRITVTLSE-GPISAPNPTLGO-----DAPAGRTFGDLARVSE-----764
Db 527 VGTKASBETVLADEKNQAVKHLAINEFWQBTKAPEGYTLDTKYPVSIKKVONNE 586
QY 765 ----VIRDIRVNRV-----PSGNCNLSAAA-----BARLVGLASAYQROEKR 805
Db 587 KNAVITRDVTAKEQVIREGDFEFKAGSADGTAEFGNDLSFKVSPLEGTXEITGAEDKA 646
QY 806 VMLHGLGFP-----LLKQFHGLLPRGM-----PNSKS-----PNP 838
Db 647 TTACNEQGFQGYGFENLPYGYLLEIEA---PEGFQKITPLEIRSTFKENKDDVAKS 703
QY 839 QWFWTLLORNO-----MPADKLTHEIITIAAVKFTBEVAAFINLP--PTCIGE 888
Db 704 EYVFTIEEGQKQPKVTVVEKLTNE-----FVSLNRLMLYDLPEKBSLTS 754
QY 889 LAQFYMANLILKYODHGYLNTLTSITGARRPRDPSSVLHWIRK---VTSAAADIETQ 945
Db 755 LATWKGNGKLNLTDFTE-LVDKL-----RYNLHEIKEDWYVVAQIDVEA- 799
QY 946 AKALLEKTENLPELWTTAFTST-----HLVRAAMNORPMVLGISIKYHGA 992
Db 800 TKAAQEKDEKAPVIAETATLANKEKTGTWKLHLKLTAEQ-----VLDKSVILFNYY 853
QY 993 AGNNRVFOAGNWSGLNGKNCVPLTFDRTRRFTIACPRGGFCPTGPPSGNRETTLS 1052
Db 854 YENKVAFEAGNE-----PVA-----KDALNN 875
QY 1053 QVRGIIVSGGAWOLAIYATVRAVGAR-----AQHAFDDMLSLTDDFLARDL-- 1102
Db 876 QAQ-----TVNCTIERHVSITKAHLEDGSGTTFHGDVMDMFDVSVTHDVL 923
QY 1103 --BELHQIITQLETPMTVEGALEAVKILDE-----KITAGD---GETPNTLAFND 1149
Db 924 GSKEAFETILYALPDGTNKEIWKSGKIEHVNDKFTKTVLAEKVDTGKYPGKFTF- 982
QY 1150 SCFSPSHDTTNSV-----LNISGSNISGSTVPLKRPPEDE 1185
Db 983 -TEINVEKDGNAVGNKHNEDLKERSQTLTPKEVPTIPSTPKQE 1024

Search completed: January 30, 2004, 13:12:34
Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

WM protein - protein search, using sw model

Run on: January 30, 2004, 13:14:08 ; Search time 44 Seconds

(without alignments)

5682.099 Million cell updates/sec

itle: US-09-769-699-2

effect score: 6294

sequence: 1 MENTQKTVTGTGLGVYVA.....DELFDLGGIPKIGNITMEM 1203

coring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

searched: 789580 seqs, 207824079 residues

total number of hits satisfying chosen parameters: 789580

inimum DB seq length: 0

aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase : Published Applications 1A:**

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:**
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:**
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:**
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:**
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:**
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:**
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:**
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:**
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:**
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:**
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:**
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:**
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:**
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:**
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:**
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:**
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:**
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Length	DB ID	Description
1	3054	48.5	1452	14	US-10-050-673-2
2	3015.5	47.9	1196	12	US-10-200-562-200
3	3015.5	47.9	1196	12	US-10-237-551-200
4	3015.5	47.9	1196	12	US-10-237-551-232
5	788	12.5	274	12	US-10-237-551-231
6	736	11.7	248	12	US-10-237-551-230
7	123.5	2.0	850	15	US-10-156-761-9121
8	120.5	1.9	2597	10	US-09-905-129-2
9	120.5	1.9	2597	10	US-09-905-129-10
10	120.5	1.9	2597	10	US-09-905-129-13
11	120.5	1.9	2597	10	US-09-905-129-13
12	120.5	1.9	2597	10	US-09-905-129-13
13	120.5	1.9	2597	10	US-09-905-129-13
14	120	1.9	5245	12	US-10-329-079-11
15	118	1.9	19608	12	US-10-084-846A-8

16	114.5	1.8	1496	12	US-10-021-660-125	Sequence 125, App
17	114.5	1.8	1496	12	US-10-331-456A-28	Sequence 28, Appl
18	113.5	1.8	1477	12	US-10-274-583-20	Sequence 20, Appl
19	112	1.8	6620	12	US-10-080-334-290	Sequence 290, App
20	111	1.8	5215	10	US-09-861-289-2	Sequence 2, Appl
21	111	1.8	5215	10	US-09-860-846-2	Sequence 2, Appl
22	111	1.8	5215	11	US-09-988-846B-2	Sequence 2, Appl
23	111	1.8	5215	11	US-09-836-821-2	Sequence 2, Appl
24	111	1.8	5215	12	US-10-271-889-45	Sequence 45, Appl
25	110.5	1.8	3069	10	US-09-712-363-246	Sequence 246, App
26	109.5	1.7	2630	14	US-10-077-130-2	Sequence 2, Appl
27	109.5	1.7	7968	14	US-10-077-130-5	Sequence 5, Appl
28	108	1.7	499	12	US-10-369-493-10577	Sequence 10577, A
29	108	1.7	2596	15	US-10-307-019-6	Sequence 6, Appl
30	107.5	1.7	8026	12	US-10-132-134-12	Sequence 12, Appl
31	107	1.7	905	12	US-10-369-493-2550	Sequence 2550, A
32	106.5	1.7	1299	15	US-10-156-761-12449	Sequence 12449, A
33	106.5	1.7	3298	12	US-10-174-677-8	Sequence 8, Appl
34	106.5	1.7	3298	12	US-10-120-801-51	Sequence 51, Appl
35	106.5	1.7	3298	15	US-10-160-758-16	Sequence 16, Appl
36	106.5	1.7	707	15	US-10-156-761-9979	Sequence 9979, Ap
37	105.5	1.7	1563	12	US-10-334-443-34	Sequence 34, Appl
38	105.5	1.7	1967	15	US-10-219-834-85	Sequence 85, Appl
39	105.5	1.7	1967	15	US-10-225-567A-575	Sequence 575, App
40	104.5	1.7	635	11	US-09-851-847-5	Sequence 5, Appl
41	104.5	1.7	1471	7	US-08-811-519A-1	Sequence 1, Appl
42	104.5	1.7	1515	12	US-10-240-154-8	Sequence 8, Appl
43	104.5	1.7	6304	15	US-10-147-026-16	Sequence 16, Appl
44	104	1.7	3192	12	US-10-132-134-10	Sequence 10, Appl
45	103	1.6	3472	12	US-10-029-120-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-10-050-673-2

; Sequence 2, Application US/10050673
; Publication NO. US20020151033A1
; GENERAL INFORMATION:
; APPLICANT: David M. Knipe
; APPLICANT: Travis J. Taylor
; APPLICANT: Elizabeth McNamee
; TITLE OF INVENTION: Replication Competent Virus Expressing A
; TITLE OF INVENTION: Fusion Protein
; FILE REFERENCE: HU98-05
; CURRENT APPLICATION NUMBER: US/10/050.673
; CURRENT FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: US/09/127.227
; PRIOR FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1452
; TYPE: PRT
; ORGANISM: herpesvirus
US-10-050-673-2

Query Match 48.5%; Score 3054; DB 14; Length 1452;
Best Local Similarity 49.8%; Pred. No. 8.3e-294;
Matches 609; Conservative 200; Mismatches 366; Indels 48; Gaps 17;

QY	1	MENTQKTVT---VPTGPIGVYV--ACRVEDLDLEISFLAARSTDSIALPLMRNLTVE	55
DB	1	METKEPTATTIKVPPGPIGVYARACSEGIEL--LALLSARSGADAVAPLVGLTVE	58
QY	56	KTFSSLLAVVGARTTGAGAGITLKTTHSHFYEVFVGGKHVLPSSAAPNLTRACNA	115
DB	59	SGFENAVVVGSRITTLGCTAVSLKUTPSHYSSVYVFGHRLDPTQAPNLTRUCER	118
QY	116	ARRPFGRCQPPVDGAVETTGAEICTRLGLEPENTILYLVVTALFKFAVFCNVFLHY	175
DB	119	ARRHFGSDYTPRFGDLKHETTGEALCERGLDPRDALLVWTEGKFAVCINNTLHL	178

09/769,699
Search Notes

176 GGLDIVHNGDVIRIPLFPVQLFMPDVRNLRVLPDPFNTHRSIGEGFVYPTPTNTGLCH 235
179 GGSOKVTIGGAHVIRIPYPLQFMPDFSRVIAFPFNHRSIGENFTYPLFPNRLNR 238
236 LIHCVIAPMAVALVRNVTAVAGAHLAFDENHGAVALPPDITYTFOSSSGTATTAR 295
239 LLFEAVGPAVALRCRNDVAVARAHLAFDENHGAVALPADITTFAFESQO--KTPR 296
296 GARNDVNSTKPSGSGFERLASIMAAUTALHAETVINTGIEETPTDIKEMPMFIGM 355
297 GGR-----DGGKGPGAGFEORLASVAGDAALALESVMAVDEBPTDISAWPLCEGQ 351
356 EGTLPRLNALGSYARVAGVIGAMVFPNSALYTEVEDSGMTEADGGPGPFRFYQF 415
352 DTAARANAVAYLARAAGLVGMVFTNSALHUTEVDDAGPADPKDHSK-PGFYFFLV 410
416 AGPFLAANPQDRDGHVL-----SSQSTGSSNTEFSDYLALICGAPILARLLFYL 468
411 PGTHVAANPOVDRECHVVPFEGRPTAPLVGGTQ-EPAGEHLAMLGFSALLAKMLFYL 469
469 EBCDAGFTGGHG-DALKYVTGTFDSEIPCLCEKHTRPVCAHTVHRLRQRMFRQAT 527
470 ERCDGGVIVGRQEMDVRYADSNQTDVPCNLCTFDRHACVHTTLMRLRARHPKFSAA 529
528 RQPIGVFTWNSQVSDCDPLGNVAPYLILKRPQDOTAAKATMODTYRATLERLFDLEQ 587
530 RGAIGVFTWNSMYSQVSDCDVLGNVAASALKR-ADGSETARTINQEIYRATERVVALET 588
588 ERLDRGAPCSGSLSSVIVDHTFTFRILDLRARIEQTTQPMKVLVTRDYKIREGLS 647
589 LQYVDQAVPTFAMGRLETTITNREALHTVNNRVQVVDREVEQLMRLNVEGRNFKPRDGLG 648
648 EATHSMALTDPDYSAGFCTINELVKRTHLAVQDLALSOCHVFCQOQVEGRNFRNQF 707
649 EANHMSLTLDPAFCPCPLQLLGRSLAVTQDLALSQCHGVFAGQSGVEGRNFRNQF 708
708 PVRRRPVDLFGNGFISTRSITVTLSEG-PVSNPNTLQGDAPAGRTFGDLARISVEVI 766
709 FVLRRRVMDNFNGFISAKTLTVALSEGAICAPSUTAGOTAPASSFEGDVARVTLGFP 768
767 RDRVRNVRVESCNTNLSEPARLVLGLASVQROKREYDMLHAGLPLKQFGLLFP 826
769 KELRVASRVLFAGASANAEEAKARVASLOSAYQKEDKEVDILLGLPLKQFHAALFP 828
827 RMPPNSKSPNQPWFLLQRNQMPADKLTHEBITTAAVKPTEEYAAINFINLPPTCI 886
829 NGKPPGSGNQPNQFWTALQRLPALLSREDIETIAFKFSLDYGAIFINLAPNV 888
887 GELAQFMANLILKYCDHSOYLINTLTSITGARRPRDPSVLHWIRKDVSAADIEQA 946
889 SELAMVYMANQILRYCDHSTYFINTLTALIAGRSRPPSVQAAAAW---SAQCGAGLEAGA 945
947 KALLEXTENPELWTTAFTSTHLVRAAMNORMVVLGISISKYHGAAGNVRVQAGWSG 1006
946 RALMDAVDAHFGANWSMFASCNLLRPVMAARVWVLGISISKYHGAAGNVRVQAGWSG 1005
1007 LGGKXVCPLFTDRTRFIIACPRGGFCPTVGPSSGNRETTLDQVRGIIVSGAMVQ 1066
1006 LMGGKXACPLIFDRTRKFLVACPRAGFYCAASNLCGGAHSESLCEQLRGIIISGGAIVA 1065
1067 LAIYATVRAVGAQAQMAFDWLSITDDEFLARDLERLHQIOTLTPMVEGAL--- 1123
1066 SSVFATVKSGLPRTQOQIEQDWLALLEDEYUSEEMELTARALGRNGEWSVDAALAVA 1125
1124 -BAVKILDEKTTAGDGETPTNLAFNFD--SCFPHSDTTSNVINTSGSNISSGTVPLKRP 1180
1126 HEAEALVSQLNAGE-----VFNFGDGC-----DDNATPGGPGAPGAPAGKGA 1173
1181 PEDDELFDLSGIPKXGNITMEM 1203
1174 FHGDOPFG-EGPPDKKGDJLDM 1195

RESULT 2
US-10-200-562-200
; Sequence 200, Application US/10200562
; Publication NO. US20030165819A1
; GENERAL INFORMATION:
; APPLICANT: McGowen, Patrick
; APPLICANT: Hosken, Nancy A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C2
; CURRENT APPLICATION NUMBER: US/10/200.562
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 200
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: HSV2
US-10-200-562-200

Query Match 47.94; Score 3015.5; DB 12; Length 1196;
Best Local Similarity 49.04; Pred No. 4e-290;
Matches 595; Conservative 213; Mismatches 366; Indels 41; Gaps 16;
QY 4 TQKTVPTGTPGLGVY--ACRVEDLDLEISFLAARSTSDIALPLMRLNLTVEKFTSS 61
DB 7 TTTTVKVPFGPMGVYGRACPAEGLEL--LSLLSARSGDADVAVALIIVGLTVESGFEAN 64
QY 62 LAVVSGARTTGLAGATTLLKTTSHFVPSVVFPHGKHVLPSSAAPNLACNAARERG 121
DB 65 VAAVVGSTTGLGTAVSLKMLSHVSPSVVPHGGRHAPSTQAPNLTLCERARPHFG 124
QY 122 PSRCQBPVGVAVETTGAEICTRLGLEPENTILYLVVLTALFKEAVMCNVFLHYGLDIV 181
DB 125 FADYAPRPCDLKHEHTGDALCERLGLDPRALYLVITEGFEAVCISNTEHLGMDKV 184
QY 182 HINHGDIIRPLFPVQLFMPDVRNLRVLPDPFNTHRSIGEGFVYPTPTNTGLCHLHDCV 241
DB 185 TIGDAEVHRIPTVPLQMFMPDFSRVIAFPFNHRSIGENFTYPLFPNRLNR 244
QY 242 IAPMAVALVRNVTAVAGAHLAFDENHGAVALPPDITYTFOSSSGTATTARGARRND 301
DB 245 VGPAVALRARNVDVAAAAHLAFDENHGAALPADITTFAFESQO--KPGQAR--- 299
QY 302 VNSTKPSGSGFERLASIMAAUTALHAETVINTGIEETPTDIKEMPMFIGMCTLPR 361
DB 300 --DAGNKGFGAGFEORLASVAGDAALALESVMAVDEBPTDISAWPLCEGQ 475
QY 362 LNALGSYARVAGVIGAMVFPNSALYTEVEDSGMTEADGGPGPFRFYQFAGPHLA 421
DB 358 AGAVGAYLARAAGLVGMVFTNSALHUTEVDDAGPADPKDHSK-PSFYRFFLVGTHVA 416
QY 422 ANPQTRDQGHVL-----SSQSTGSSNTEFSDYLALICGAPILARLLFYLERCDA 474
DB 417 ANPQDREGHVVPFEGRPTAPLVGGTQ-EPAGEHLAMLGFSALLAKMLFYLERCDDG 475
QY 475 AFTGGHG-DALKYVTGTFDSEIPCLCEKHTRPVCAHTVHRLRQRMFRQATPQF 533
DB 476 VIVGRQEMDVFRVADSGQTDVPCNLCTFDRHACVHTTLMRLRARHPKFSAA 535
QY 534 FGTWNSQVSDCDPLGNVAPYLILKRPQDOTAAKATMODTYRATLERLFDLEQERLDR 593
DB 536 FGTWNSQVSDCDPLGNVAPVSAFKR-ADGSENTRTINGSTYRAATERVAALEALQYVDQ 594
QY 594 GAPSSEGLSSVIVDHTFTFRILDLRARIEQTTQPMKVLVTRDYKIREGLSEATHSM 653
DB 595 APTALGRLETTIGNREALHTVNNIKOLVREVEQLMRLNTEGRNFKPRDGLAEANHAM 654
QY 654 ALTFDPYSAGFCPTINFLVNRTHLAVQDLALSOCHVFCQOQVEGRNFRNQFVPLRRR 713
DB 655 SLSDPYTTCGPCPLQLLARRSNLAVQDLALSOCHVFCQOQVEGRNFRNQFVPLRRR 714

QY	714	FVDLENGGFISTRITVTLSSEG-PVSAPNPTLGDAPAGRTFGDILARVSVSEVIRDKV	772
Db	715	VMDLFNNGLSAKTLTUTVALSAAICAPSLTAGTAPAESFEGDVAVRVLTPFKELRVK	774
QY	773	NRVFSGNCTNLSEAARALVGLASAYOROEKRVDMLHGLGFLKFHLLPRGWPPN	832
Db	775	SRVLFAAGASANAASEAKARVASLOSAYQKPKXVDILLGPLGFLLQFHAVIFPNKPPG	834
QY	833	SXSPNQFWLLOQNQPADKLTHEIITIAVKRTEBYAINTINLPPTTCIGELAOF	892
Db	835	SNQPNPWFALQONCLPARLLSGREDIETIAFKRPISLDYGAINFINLAPNVSELAMY	894
QY	893	YMANLILKYCDHSOVLNTLTSIIITGARRRDRSSVLHWIKDVTSAADIETOAKALLEK	952
Db	895	YMANOLLYCDHYFYFNITLITAVTAGRRPPSVQAAAAPQ---GGAGLEAGARALMDS	951
QY	953	TENPELMTTAFTSHLVRAAMNQRPMVLGISISKYHGGAANNRVFOAGNWSGLNGKN	1012
Db	952	LDAHPGAWTSMFASCNLLRPVMAARPVMVLGLSISKYGMAGNDRVFOAGNASLGGKN	1011
QY	1013	VCPLTFEDTEREFTIACPRGGFCVTPGPSSGNRETTLSDOVRGIIVSGAMVQLAIAT	1072
Db	1012	ACPILLIDTRTKFVLACPRAGFVCAASSLGGGAHEHSLCEQLRGITIAEGGAIVASSVFVA	1071
QY	1073	VYRVARQAQMADFDDWLSITDDDEFLARDLEELEDQIOIOTLETPTWVEGAL--- <td>1128</td>	1128
Db	1072	TVKSIGPRTQQLIEDWLALDEDYLSEEMMETTRALERHGEHSTDAALEVAHEAEAL	1131
QY	1129	LDEKTAGDGETPTNLAFNPDCBPESHDTNSVNLISGNSISTVPGLKRPDPDELFD	1188
Db	1132	VSQJGAAGE-----VFNFGDFGEDDHAAESFGGLAAA--AGAAGVARKRAFHDGDPFG	1182
QY	1189	LSGIPIKKHGNTMEM	1203
Db	1183	-EGPEKK-DLTDLM	1195
RESULT 5			
US-10-237-551-231			
; Sequence 231, Application US/10237551			
; Publication No. US20030165820A1			
; GENERAL INFORMATION:			
; APPLICANT: Day, Craig H.			
; APPLICANT: Hosken, Nancy A.			
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND			
; FILE REFERENCE: 210121.538C3			
; CURRENT APPLICATION NUMBER: US/10/237,551			
; NUMBER OF SEQ ID NOS: 254			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 231			
; LENGTH: 274			
; TYPE: PRT			
; ORGANISM: Herpes simplex virus			
US-10-237-551-231			
Query Match 12.5%; Score 788; DB 12; Length 274;			
Best Local Similarity 56.5%; Pred. No. 2.6e-69;			
Matches 153; Conservative 41; Mismatches 75; Indels 2; Gaps 1;			
QY	28	LEBISFLAARSTDALLPLMRNLTYEKTFTSSLAUVVGARTTGLAGAGITLKLTSHF	87
Db	2	LELSLSARSODADVAVAPLIVGLTVESFEANVAVVGSRTTGLGTAVSLKMPSHY	61
QY	88	YPVSFVFGHKVLPSSAAPNLTRACNAARERFGRSQCPGVDPVGVATTGAEICTRLGL	147
Db	62	SPSVYVFGHRHLASTQAPNLTRLCERARHFESDYAPPCDLKHETTGDALCELGL	121
QY	148	EPENTLXLVLTALKPAEVCMCNVFLHYGGLDIVHINHGDVIRIPLFPVOLFMFDVNRLV	207

Db 122 DDPRLALYLITEGFEAREAVCSNTFLHLGGMDKVTIGDAEVHRIPIYVPLQMFPPDSRVI 181
 QY 208 PDPFNTHRSIGEGFYVPPFYNTGLCHLHDCVIAPMAVALRVNVTAVAGAAHLAPD 267
 Db 182 ADFFNCNHSIGENFYPLPFNFNRLPRLILFEAVGFAAVALRARNVDVAVAAAAHLAPD 241
 QY 268 ENHEGAVLPDPITYTYFQSSSSGTTTARGAR 298
 Db 242 ENHEGAALPADITTTAFASQG--KFORGAR 270

RESULT 6

S-10-237-551-230

Sequence 230, Application US/10237551

Publication No. US20030165820A1

GENERAL INFORMATION:

APPLICANT: Day, Craig H.

APPLICANT: Hosken, Nancy A.

APPLICANT: Parsons, Joseph M.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND

FILE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION

FILE REFERENCE: 210121.538C3

CURRENT APPLICATION NUMBER: US/10/237,551

CURRENT FILING DATE: 2002-09-06

NUMBER OF SEQ ID NOS: 254

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 230

LENGTH: 248

TYPE: PR

ORGANISM: Herpes simplex virus

S-10-237-551-230

Query Match 11.7%; Score 736; DB 12; Length 248;
 Best Local Similarity 57.7%; Pred. No. 3.3e-64;
 Matches 139; Conservative 36; Mismatches 66; Indels 0; Gaps 0;

Y 47 PLMRNLTVKFTTSSLAUVSGARTTGLAGAGITLKLTTSHFPYPSVVFHGGKHVLPSSAA 106
 b 3 PLVGLTVESGFEANVAAVGSRITGLGTAVALSLKMLPSHYSVSVVFGGRLHAPSTQA 62
 Y 107 PNLTACNAARERFCFSRCQPPVDGAVETTGAGICTFGLGLEPENTILYLVVTFALFKEAV 166
 b 63 PNLTLCERARRHFGFSYARPCDLKHETTGDCALCERGLDPPRALYLVITGFEAV 122
 Y 167 FMCNVLHYGLDIVHNGDVIRIPLFPVQLFMPDVNRLVPDPFNTHRSIGEGFYPT 226
 b 123 CISNTFLHLGGMDKVTIGDAEVHRIPIYVPLQMFPPDSRVIADFFNCNHSIGENFYPL 182
 Y 227 PFYNTGLCHLHDCVIAPMAVALRVNVTAVAGAAHLAFDENHEGAVLPDPITYTYFQS 286
 b 193 PFFNRPLARLLFEAVVGAVALRARNVDVAVAAAAHLAFDENHEGAALPADITTTAFEA 242
 Y 287 S 287
 b 243 S 243

RESULT 7

S-10-156-761-9121

Sequence 9121, Application US/10156761

Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 9121

; LENGTH: 850

; TYPE: PR

; ORGANISM: Streptomyces avermitilis

US-10-156-761-9121

Query Match 2.0%; Score 123.5; DB 15; Length 850;
 Best Local Similarity 20.5%; Pred. No. 0.031;
 Matches 182; Conservative 104; Mismatches 327; Indels 273; Gaps 37;

QY 305 TSKPSPSGGFE-----RRLASIMAAADTALHAENVIENTGIYEETPTDIDKEWPMFIM 355
 Db 3 TSRTTSPAGAEPPSRPPVGRRAHAQPADE-----DTGTDVTVAASADSVPVAGR 54
 QY 356 EGTLP-----LNALGSYARVAGVIGAMVSPNSALYLTVEVSGMTE 399
 Db 55 WGVPRPTVRAKIVCLLMVPVSLALWAY-----ATVSTAQDVARLRQVQVDTTV 105
 QY 400 AKDGGPGPSFNRFYQFAGPHLAANPOTDSDGHVLSQSTGSSNTFSVDYLAL-----ICG 455
 Db 106 RAPVAAVAALQAERAAAVRHVIDFSAEPD-----SGFRTLAARTDRAVDKURLGGHTVA 161
 QY 456 FGAPLLARLLFYLERCDAGFTGGHGDALKYVTGTFDSEIPCSCKEHTRPVCAHTTVHR 515
 Db 162 DGADLPAQVGRLETFVSGA-----EQLRSRUG-----AVLER 194
 QY 516 LRQENPRFGCATQPIGVFGTWN-----SQYSDCDPLGNAYAPYLILRKPQOTENAAKTMQDT 573
 Db 195 RARWDETFQYTRTTAAAFVGGALGTIGQAD-LGSDARVLL-----BFSRAGEALAQED 248
 QY 574 YRATLERLFDIEQERL-LDRGAPCSSEGLS-SVIVDHTFERRILDTLRARIEQTTQFM 631
 Db 249 AVLSSARLAGTLDGERLALFTGAVDTRTLTDSAVD-----LSERERAAVQGVATORA 302
 QY 632 KVLVETRDYKI---REGLESEATHSMALTFDP-----YSGAPCPPTN 669
 Db 303 YADVRTAEKVLNAPGARITAAAPQATWPAHARVQSGMRTIEADAGRVADRADPLTR 362
 QY 670 FLVKRTHLAVQDLALSOCHCVF-----YGOQVEGNFRNQFQVLRERFVD-----LFNG 720
 Db 363 GLLTAGAAVLFGLAAVAASLIVSVRIGRGLVIELISURNALAIARKLPQAMRKLRAG 422
 QY 721 GFISTRSITVTLSEGPVSAP-NPTLGDAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSG 779
 Db 423 EEIDVR-----AEAPPGPFAEDETGQVAEALSTVHRAALRAAVE-----461
 QY 780 NCTNLSEARARLVCLASAYQOEKVDMLHGALGFLKKQPHGLLPRGMPNKSXPNQ 839
 Db 462 -----RAELASGISGV-----472
 QY 840 WFTWLLQONMPADKLTHEEITITAAVKRFTVEEVAINFINLPPTCIGELAQFYMANLIL 899
 Db 473 -FVNLAERSQI---LVERQLSLDSDMERSED-----PNELSDL--FLDLHLLT 515
 QY 900 KYCDHSQVILNTLSITGA---RRPRDPSVLHWIRKDVTSAAD---IETQ-----945
 Db 516 RMRRHAESLI-----ILSGAAPGRAWMPVSLTNVWRAAVSEVEDYARVEVRQLPEASVV 570
 QY 946 AKALIEKTEENLPWLTTA--FTSTHLVRAANQRP-----MVLGISISKYHGAAGN 995
 Db 571 GAAVADLTHLMAEIVEENAAQSPPH-TRVTVGTGPGVNGYAVEVEDGLGNGKETLEBAN 629
 QY 996 NRVFOAGNWSGLNGKNCVCLPFTFDRTRRTIIA-----CPRGGFICPVTPG 1041
 Db 630 RRIEQS-----EALDLFSDRLGLFWVSLAARHGKIVHLRTSPYGGTTAVVLLP 679
 QY 1042 S-----SGNRETT--LSQVQRGIIYVSGAMVQLAIYATVAVANGARA 1081

Db 680 TALLHSGTAERVPRAADTGRD-AEPAYARVAASHQSVQAVGRPA 724

RESULT 8

US-09-905-129-2
; Sequence 2, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 540579-2007.2
; CURRENT APPLICATION NUMBER: US/09/905,129
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/207,821
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/084,944
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 60/085,673
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2

LENGTH: 2597
TYPE: PRT
ORGANISM: rattus species
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(2597)
OTHER INFORMATION: 'x' can be any amino acid

US-09-905-129-2

Query Match
Best Local Similarity 1.9%; Score 120.5; DB 10; Length 2597;
Matches 145; Conservative 113; Mismatches 287; Indels 233; Gaps 40;

2Y 416 AGPFLAANPQTD---RDGHVLSQSQTGSNTE-----FSDVYLALICG 455
Db 274 SGFLCTKTIDPSLKSKSLVTQEDNGSASTSPQDIEPFGSLNMTXXSGNKADMVCS 333
2Y 456 FGAPLLARLLFYLERCDAGFTGCHGDALKYVTGTFDSEIPCSLCEKHTRPV---CAHT 511
Db 334 IQKP-----SRTSPATFEENDYIM--LNASFSTNLVCSVDYNIHQVPWQLLAYS 382
2Y 512 TVHRLRQNP-----RQCATROPIGVGTGMSQYSDCDPLGNYPYLILRKPGD 561
Db 383 DGPLILRKPQTETPSLSRYKQVALRPEDIFTSIADVR-ADPFWFQKEKIVLQLNRT 441
2Y 562 QTEAAKATMQDTYRATLERLFDLEQERL-----LDR-----GAPCSSEG 601
Db 442 ATTLSTLQIOFSTDAQIALPRAEMPAERLKWMTILMNNPKLERTVLVGGTIALSCPGK 501
2Y 602 -----LSSVIVDHPFTR-----RILDTLRARIE-QTTTFMKVL---VETRD----- 639
Db 502 DSPHLEWLLADGSKVRAPYVSEDEGRILIDKNGKLEQWADSFAGLVHCISTNDADAV 561
2Y 640 --YKIR--EGLSEATHSMALTDPPYSGAF---CPITFLVKTHTLAVQDIALSQCHCV 691
Db 562 LTYRITVPEYGESTHDSGVQHTVVTGETLDPLCLSTG-----VPDASIS---WI 608
2Y 692 FYGQVQEGNFRNQFPVLRFRFVDFLNGGFTSITRSITVTLSSEGVPS--APNPT----- 743
Db 609 LFGNTVFSQPSRDR-----QILNNGTLRILOVT-PKQGHYQCVANPAGADPSS 657
2Y 744 -----LGODAPAGRTFDCGLARVSEVIRDIRVRKRVNRFVSGNCTNLSEAPAR 791
Db 658 FKVSQKGGQRMVHEDRAGSGGLGE-FNSVSLKQFASLK-----LSASALTGSEAGKQ- 711
2Y 792 LVGLASAYORQEKRVDMHLHGALG-FLKQFHGCLLFRGMPNKSXPQPFWTLQ---R 847
b 712 ----VSGVHRKXKHEDLHRRRGDSTLRFRE--HRQLPLSARRIDPQWAALEKAKK 765

QY 848 NOMPADKLTHEBITT-----IAAVKRFTEBYAAINFILPPTCIGELAQFYMANLILK 900
Db 766 NSVP-----KKQENTTVKVPVPLAVPLVELTDEKDSAGMI--PP----- 802
QY 901 YCDHSQYLLINTLSITGARRPRDPSSVLHWIRKDVTSAAIDTQAKALLEKTENLPELM 960
Db 803 --DEEFWLKTKASGVPRSPADSGPVNHHGFMFTSIASGTEVSTVNPQTLO-SEHLPDFK 859
QY 961 TTAFTSTHLVRAAM-----NORPMVLGISISKYHGAAGNNRVFOAGNWSG-- 1006
Db 860 LPSVINGTAVTKSNMPSIASKIEDTTNQNPFIIP-SVAEIRDSA-----QAGRASSQS 912
QY 1007 ----LNGKKNVCPLFTFDRTRRPIIACPRGGFI-----CPVTG---PSSGN 1045
Db 913 AHPVTGGN----MATYGTNTYTSFTSKASTVLOPINTPESYGPQIPIITGVSRDSSD 966

RESULT 9

US-09-905-129-10
; Sequence 10, Application US/09905129
; Patent No. US20020137705A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al

; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
; FILE REFERENCE: 540579-2007.2
; CURRENT APPLICATION NUMBER: US/09/905,129
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/207,821
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/084,944
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 60/085,673
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
LENGTH: 2597
TYPE: PRT
ORGANISM: Rattus sp.
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(2597)
OTHER INFORMATION: 'x' can be any amino acid
US-09-905-129-10

Query Match
Best Local Similarity 1.9%; Score 120.5; DB 10; Length 2597;
Matches 145; Conservative 113; Mismatches 287; Indels 233; Gaps 40;

QY 416 AGPFLAANPQTD---RDGHVLSQSQTGSNTE-----FSDVYLALICG 455
Db 274 SGFLCTKTIDPSLKSKSLVTQEDNGSASTSPQDIEPFGSLNMTXXSGNKADMVCS 333
QY 456 FGAPLLARLLFYLERCDAGFTGCHGDALKYVTGTFDSEIPCSLCEKHTRPV---CAHT 511
Db 334 IQKP-----SRTSPATFEENDYIM--LNASFSTNLVCSVDYNIHQVPWQLLAYS 382
QY 512 TVHRLRQNP-----RQCATROPIGVGTGMSQYSDCDPLGNYPYLILRKPGD 561
Db 383 DGPLILRKPQTETPSLSRYKQVALRPEDIFTSIADVR-ADPFWFQKEKIVLQLNRT 441
QY 562 QTEAAKATMQDTYRATLERLFDLEQERL-----LDR-----GAPCSSEG 601
Db 442 ATTLSTLQIOFSTDAQIALPRAEMPAERLKWMTILMNNPKLERTVLVGGTIALSCPGK 501
QY 602 -----LSSVIVDHPFTR-----RILDTLRARIE-QTTTFMKVL---VETRD----- 639
Db 502 DSPHLEWLLADGSKVRAPYVSEDEGRILIDKNGKLEQWADSFAGLVHCISTNDADAV 561

```

2Y 640 --YKIR--EGLSEATHSMALTFDPYSGAF-----CPITNPLVKRTHLAVVQDLALSQCHCV 691
Db 640 --YKIR--EGLSEATHSMALTFDPYSGAF-----CPITNPLVKRTHLAVVQDLALSQCHCV 691
2Y 562 LTYRITVVEFYGESTHDSGVQHTVVTGETLDPCLSTG-----VPDASIS---WI 608
Db 562 LTYRITVVEFYGESTHDSGVQHTVVTGETLDPCLSTG-----VPDASIS---WI 608
2Y 692 FYGQOVGEGRNFNQFQVLRREFVDFLFGGFISTRSITVTLSEGPVS---ANPT-----743
Db 692 FYGQOVGEGRNFNQFQVLRREFVDFLFGGFISTRSITVTLSEGPVS---ANPT-----743
2Y 609 LFGNTVFSQPSDR-----OILNNGTILRILOVT--PKDQGHYQCVAAANPSGADPSS 657
Db 609 LFGNTVFSQPSDR-----OILNNGTILRILOVT--PKDQGHYQCVAAANPSGADPSS 657
2Y 744 -----LGODAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCNTLSEAAAR 791
Db 744 -----LGODAPAGRTFDGDLARVSVEVIRDIRVKNRVVFSGNCNTLSEAAAR 791
2Y 658 FKVSQKQGMVEHREAGGSGLGE--PNSVSLKQPASLK-----LSASALTGSEAGKQ--711
Db 658 FKVSQKQGMVEHREAGGSGLGE--PNSVSLKQPASLK-----LSASALTGSEAGKQ--711
2Y 792 LVGLASAYQROEKRVDMHLGALG--FLKQFHGLLFPKGMPPNSKSPNPQWFWTLQ---R 847
Db 792 LVGLASAYQROEKRVDMHLGALG--FLKQFHGLLFPKGMPPNSKSPNPQWFWTLQ---R 847
2Y 712 ----VSGVHRKNKHRLHRRRGDSTLRFRF--HRRQLPLSARRIDPQWAAALLERAKK 765
Db 712 ----VSGVHRKNKHRLHRRRGDSTLRFRF--HRRQLPLSARRIDPQWAAALLERAKK 765
2Y 848 NOMPADKLTHEEIT-----IAAVKRFTEEYAAINFILNPPCTICIGELAQFYMANILK 900
Db 848 NOMPADKLTHEEIT-----IAAVKRFTEEYAAINFILNPPCTICIGELAQFYMANILK 900
2Y 766 NSVP-----KKQENTIVKPVPLAVPLVELTDEEKASGM---PP-----802
Db 766 NSVP-----KKQENTIVKPVPLAVPLVELTDEEKASGM---PP-----802
2Y 901 YCDHSQYLINTLSITGARRPRDPSSVLHWIRKDVTSAADIEFOAKALLEKTENLPELW 960
Db 901 YCDHSQYLINTLSITGARRPRDPSSVLHWIRKDVTSAADIEFOAKALLEKTENLPELW 960
2Y 803 --DEEFVWLKTKASGVPGRSPTADSGPVNHGFMTSIASGTEVSTVNPOTLQ--SEHLPDFK 859
Db 803 --DEEFVWLKTKASGVPGRSPTADSGPVNHGFMTSIASGTEVSTVNPOTLQ--SEHLPDFK 859
2Y 961 TTAFTSTHLVRAAM-----NORPMVVLGISISKVHGAAGNNRVFOAGNWSG--1006
Db 961 TTAFTSTHLVRAAM-----NORPMVVLGISISKVHGAAGNNRVFOAGNWSG--1006
2Y 860 LFSYNTGTAVTKSNPSTASKIEDTTNQNFIIIFP--SVAEIRDSA-----QAGRASSQS 912
Db 860 LFSYNTGTAVTKSNPSTASKIEDTTNQNFIIIFP--SVAEIRDSA-----QAGRASSQS 912
2Y 1007 ---LNGGKNVCPLFTFDRTRFIIACPRGGFI-----CPVTG---PSSGN 1045
Db 1007 ---LNGGKNVCPLFTFDRTRFIIACPRGGFI-----CPVTG---PSSGN 1045
2Y 913 AHPVTGNN---MATYGHNTYSSFTSKASTVLPQINPTESYGPQIPITGVSRPSSSD 966
Db 913 AHPVTGNN---MATYGHNTYSSFTSKASTVLPQINPTESYGPQIPITGVSRPSSSD 966

RESULT 11
US-09-991-630-2
; Sequence 2, Application US/09991630
; Patent No. US200201514A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THEREOF
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 540579-2007.3
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US/09/991,630
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 09/729,485
; PRIOR FILING DATE: 2000-12-04
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 2597
; TYPE: PRT
; ORGANISM: Rattus species
; FEATURE:
; NAME/KEY: misc_feature

```

```

Query Match
Best Local Similarity 18.6%; Pred. No. 0.46; Length 2597;
Matches 145; Conservative 113; Mismatches 287; Indels 233; Gaps 40;

416 AGPHLAANPOTD---RDGHVLSQSTGSSNTE-----FSDVYLALICG 455
274 SGAFLLKTLFLLERCDAGFTGGHGDALKYVTGTFDSEIPCSLCEKTRPV---CAHT 511
456 FGAPLLARLLFLLERCDAGFTGGHGDALKYVTGTFDSEIPCSLCEKTRPV---CAHT 511

```

RESULT 10

```

3-09-905-129-13
Sequence 13, Application US/09905129
Patent No. US20020137705A1
GENERAL INFORMATION:
APPLICANT: Einat, et al
TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THEREOF
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: 540579-2007.2
CURRENT FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US/09/905,129
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 60/207,821
PRIOR FILING DATE: 2000-05-30
PRIOR APPLICATION NUMBER: 60/084,944
PRIOR FILING DATE: 1998-05-11
PRIOR APPLICATION NUMBER: 60/085,673
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patent in version 3.0
; SEQ ID NO 13
; LENGTH: 2597
; TYPE: PRT
; ORGANISM: Rattus sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(2597)
; OTHER INFORMATION: 'x' can be any amino acid
; -09-905-129-13

```


LOCATION: (1)...(2597)
OTHER INFORMATION: 'x' can be any amino acid
US-09-991-630-2

Query Match 1.9%, Score 120.5; DB 10; Length 2597;
Best Local Similarity 18.6%; Pred. No. 0.46;
Matches 145; Conservative 113; Mismatches 287; Indels 233; Gaps 40;

QY 416 AGHLAANPQD---RDGHVLSQSTGSSNTE-----FSDVYLALIG 455
DB 274 SGAFCTKPTIDPSLKSLSVTQEDNGSASTSPQDIEPFGSLNMTXXSGNKADMVCS 333
QY 456 FGAPLLARLLFYLERCDAGTGGHGDALKYVTGTFDSEIPCSLCEKHTRPV-----CAHT 511
DB 334 IQKP-----SRISPTAFTENDYIM--LNASTNLVCSVDYHNIQPVWQLLALYS 382
QY 512 TVHRLRQMP-----RFGQATROPICVGTWNSQYSDCPGLNAYAPYLILKPGD 561
DB 383 DSEPLIERKQLTETPSLSRYKQVALRPEDIFTSIADVR--ADPFWQOEKIVLQJLNT 441
QY 562 QTEAAKATMDTYRATLERLFDLEQERL-----LDR-----GAPCSSEG 601
DB 442 ATTLSLTIQIFSTDAQIALPRAEMRAERLKWMLMNNPKLERTVLVGGTIALSCPGKG 501
QY 602 -----LSSVIVDHPTFR-----RILDTLRARIE--QTTTFMKVL-----VETRD----- 639
DB 502 DPSPHLEWLLADGSKVRAPYVSEDGRLDKNGKLEOMADSFDAGLYHCISTNDADADV 561
QY 640 --YKIR--EGLSEATHSMALTDPYSGAF---CPITNVLKTRTHLAVVQDLALSOCHVCV 691
DB 562 LTVRIITVPEYGESTHDSGVQHTVVTGETLDPCLSTG-----VPDASIS---WI 608
QY 692 FYCQOQVEGNFRNQFVLRFRFVDFNGGFTSTRSITVTLSEGPVS--APNPT----- 743
DB 609 LPGNTVFSQPSRDR-----QILNNGTILRILOVT--PKDQGHYQCVAAANPSGADFS 657
QY 744 -----LGODAPAGRTFDGLARVSVEVIRDIRVKNRVVFSNGCTNLSAARAR 791
DB 658 FKVSQKQGVQVVEHREAGGSLGE--PNSSVSLKQPASLK-----LSASALTGSEAGKQ- 711
QY 792 LVGLASAYORQEKRVDMHLGALG--FLKQFHGLLPPRGMPNPKSPNPFQWFTLIQ---R 847
DB 712 ----VSGVHKXKHRLIHRRRGDSTLRFRF--HRRQLPUSARRIDPQWMALEKAKK 765
QY 848 NOMPADKLTHEIIT-----IAAVKRTTEYAAINFNLPTTCIGELAOFYMANLILK 900
DB 766 NSVP---KKQENTTVKPVPLAVPLVELTDEKDGSMI--PP----- 802
QY 901 YCDHSQYLINTLTSITGARRPRDPSVLEHWIRKDVTSAAADIETQAKALLEKTENLPELM 960
DB 803 --DEEFWLKTKASGVPGSRPTADSGPVNHGFWMTSIAGTEVSTVNPOTLQ--SEHLPDFK 859
QY 961 TTAFTSTHLVRAAM-----NORPMVVLGISISKYHGAAGNNRVFQAGNWSG-- 1006
DB 860 LFSVINGTAVTKSMNPSIASKIEDITNQNPILIFF--SVAEIRDSA-----QAGRASSQS 912
QY 1007 ---LNGGKVCVLPFTDRTRRIIACPRGGFI-----CPVTG---PSSGN 1045
DB 913 AHPVTGCGN---MATYGHNTYSSFTSKASTVLQPINPTESYGVQPIPTIGVSRPSSD 966

RESULT 12
US-09-991-630-10
Sequence 10, Application US/09991630
Patent No. US200201515141
GENERAL INFORMATION:
APPLICANT: Elnat, et al
TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
FILE REFERENCE: AND USES THEREOF
FILE REFERENCE: 540579-2007.3
CURRENT APPLICATION NUMBER: US/09/991,630
CURRENT FILING DATE: 2001-11-06
PRIOR APPLICATION NUMBER: 09/905,129

PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: 09/802,318
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 09/729,485
PRIOR FILING DATE: 2000-12-04
NUMBER OF SEQ ID NOS: 28
SOFTWARE: Patent in version 3.0
SEQ ID NO 10
LENGTH: 2597
TYPE: PRT
ORGANISM: Rattus species
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(2597)
OTHER INFORMATION: 'x' can be any amino acid
US-09-991-630-10

Query Match 1.9%, Score 120.5; DB 10; Length 2597;
Best Local Similarity 18.6%; Pred. No. 0.46;
Matches 145; Conservative 113; Mismatches 287; Indels 233; Gaps 40;

QY 416 AGHLAANPQD---RDGHVLSQSTGSSNTE-----FSDVYLALIG 455
DB 274 SGAFCTKPTIDPSLKSLSVTQEDNGSASTSPQDIEPFGSLNMTXXSGNKADMVCS 333
QY 456 FGAPLLARLLFYLERCDAGTGGHGDALKYVTGTFDSEIPCSLCEKHTRPV-----CAHT 511
DB 334 IQKP-----SRISPTAFTENDYIM--LNASTNLVCSVDYHNIQPVWQLLALYS 382
QY 512 TVHRLRQMP-----RFGQATROPICVGTWNSQYSDCPGLNAYAPYLILKPGD 561
DB 383 DSEPLIERKQLTETPSLSRYKQVALRPEDIFTSIADVR--ADPFWQOEKIVLQJLNT 441
QY 562 QTEAAKATMDTYRATLERLFDLEQERL-----LDR-----GAPCSSEG 601
DB 442 ATTLSLTIQIFSTDAQIALPRAEMRAERLKWMLMNNPKLERTVLVGGTIALSCPGKG 501
QY 602 -----LSSVIVDHPTFR-----RILDTLRARIE--QTTTFMKVL-----VETRD----- 639
DB 502 DPSPHLEWLLADGSKVRAPYVSEDGRLDKNGKLEOMADSFDAGLYHCISTNDADADV 561
QY 640 --YKIR--EGLSEATHSMALTDPYSGAF---CPITNVLKTRTHLAVVQDLALSOCHVCV 691
DB 562 LTVRIITVPEYGESTHDSGVQHTVVTGETLDPCLSTG-----VPDASIS---WI 608
QY 692 FYCQOQVEGNFRNQFVLRFRFVDFNGGFTSTRSITVTLSEGPVS--APNPT----- 743
DB 609 LPGNTVFSQPSRDR-----QILNNGTILRILOVT--PKDQGHYQCVAAANPSGADFS 657
QY 744 -----LGODAPAGRTFDGLARVSVEVIRDIRVKNRVVFSNGCTNLSAARAR 791
DB 658 FKVSQKQGVQVVEHREAGGSLGE--PNSSVSLKQPASLK-----LSASALTGSEAGKQ- 711
QY 792 LVGLASAYORQEKRVDMHLGALG--FLKQFHGLLPPRGMPNPKSPNPFQWFTLIQ---R 847
DB 712 ----VSGVHKXKHRLIHRRRGDSTLRFRF--HRRQLPUSARRIDPQWMALEKAKK 765
QY 848 NOMPADKLTHEIIT-----IAAVKRTTEYAAINFNLPTTCIGELAOFYMANLILK 900
DB 766 NSVP---KKQENTTVKPVPLAVPLVELTDEKDGSMI--PP----- 802
QY 901 YCDHSQYLINTLTSITGARRPRDPSVLEHWIRKDVTSAAADIETQAKALLEKTENLPELM 960
DB 803 --DEEFWLKTKASGVPGSRPTADSGPVNHGFWMTSIAGTEVSTVNPOTLQ--SEHLPDFK 859
QY 961 TTAFTSTHLVRAAM-----NORPMVVLGISISKYHGAAGNNRVFQAGNWSG-- 1006
DB 860 LFSVINGTAVTKSMNPSIASKIEDITNQNPILIFF--SVAEIRDSA-----QAGRASSQS 912
QY 1007 ---LNGGKVCVLPFTDRTRRIIACPRGGFI-----CPVTG---PSSGN 1045
DB 913 AHPVTGCGN---MATYGHNTYSSFTSKASTVLQPINPTESYGVQPIPTIGVSRPSSD 966

RESULT 13

US-09-991-630-13
; Sequence 13, Application US/09991630
; Patent No. US20020151514A1
; GENERAL INFORMATION:
; APPLICANT: Einat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 540579-2007.3
; CURRENT APPLICATION NUMBER: US/09/991,630
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: 09/905,129
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 09/729,485
; PRIOR FILING DATE: 2000-12-04
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 2597
; TYPE: PRT
; ORGANISM: Rattus species
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(2597)
; OTHER INFORMATION: 'x' can be any amino acid

S-09-991-630-13

Query Match 1.9%; Score 120.5; DB 10; Length 2597;
Best Local Similarity 18.6%; Pred. No. 0.46;
Matches 145; Conservative 113; Mismatches 28; Indels 233; Gaps 40;

Y 416 AGPHLAANPQD---RDGHVLSQSTGSNTE-----FSDVYLAICG 455
b 274 SGAFCTCTPTDPSLKSLSVTDQNGSASTSPQDFIEPFGSLNMTXXSGNKADMVCS 333
Y 456 FGAPLLARLLFYLERCDAGFTGHDGALKVVTGTFDEIPECSCEKHTRPV---CAHT 511
b 334 IQKP-----SRTSTATEENDIYM--LNASFNLVCSVDYVHIIQVWQLLALYS 382
Y 512 TVHRLRQMP-----RFGQATROPIGVGTMSQVSDCDPLGNAYPIYLIRKPGD 561
b 383 DSPILEREKPOLTETPSLSRYKQVALREFDIFTSIEADVA-ADPFWFQBEKIVLQNLRT 441
Y 562 QTEAAKATWQTYRATLERLFIDLEQERL-----LDR-----GAPCSSEG 601
b 442 ATTISTIQISTDAQIALPRAEMRAERLKTWMLMNNPKLERTVLVGGTIALSCPGRG 501
Y 602 -----LSSVIVDHTFR-----RLDTLRARIE-QTTTQFMKVL---VETRD----- 639
b 502 DPSPHLEWLLADGSKVRAPYVSEDEGRILIDKNGKLEQMAQSDPAGLYHCISTNDADV 561
Y 640 --YKIR--EGISEATHSALTDPYSGAF-----CPITNFLVKRTHLAVQDLALSQCHCV 691
b 562 LTYRITVVPEYGESTHDSGVQHTVVTGETLPLCLSTG-----VPDASIS---WI 608
Y 692 FYGOVEGRNFRNQFQVRLRRFRVLENGGFISTRTSVTLSEGPVS--APNPT----- 743
b 609 LFGNTVFSQPSRDR-----QILNNGTILRILOVT-PKQOHYOCVLANFGADFS 657
Y 744 -----LGODAPAGTFTGDLARVSEVIRDIRVKRNVFVSGNCTNLSEARAR 791
b 658 FKVSQKKGQRMVDEHREAGSGGLGE-PNSVSLSLKQFASLK-----LSASALTGSEAGKQ- 711
Y 792 LVGLASAYQOREKRVDMHLHGALG-FLKQFHGLFPRGMPNSKSPNQPFWTLIQ---R 847
b 712 ----VSGVHRKMKRDLLHRRRGSTLRRFRE--HRRQLPUSARRIDPQRWALLERAKK 765
Y 848 NOMPADKLTHEEIT-----IAAVKFTBEYAAINFILPPTCIGELAQFYMANLILK 900
b 766 NSVP-----KKQENTVTRVPVLAFLVELTDEKDSGMI--PP----- 802

QY 901 YCDHSQVILNTLSITGARRPRDPSSVLHWIRKDVTSAAIDETQAKALLEKTEMLPELM 960
Db 803 --DEEFWLKTKASGVPGSPADSGPVNKGFWTSIASGTEVSTVNPQTLO-SEHLPDFK 859
QY 961 TTAFSTHLYRAAM-----NQRPMVLGISISKYHGAAGNNRVFQAGNWSG-- 1006
Db 860 LFSVTNGTAVTKSMNPSIASKIEDTTNQNPIIFP-SVAEIRDSA-----QAGRASSQS 912
QY 1007 ---LNGGKNVCPLFTEDRTRRPIIACPRGGFI-----CPVTG---PSSGN 1045
Db 913 AHPVTGNN-----NATYGHNTWYSSFTSKASTVLQINPTESYGQPIITGVSRPSSD 966

RESULT 14
US-10-329-079-11
; Sequence 11, Application US/10329079
; Publication No. US20030198981A1
; GENERAL INFORMATION:
; APPLICANT: FARNET, Chris
; APPLICANT: ZAZOPOULOS, Emmanuel
; APPLICANT: STAFFA, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF LIPOPEPTIDES
; FILE REFERENCE: 3002-11US
; CURRENT APPLICATION NUMBER: US/10/329,079
; CURRENT FILING DATE: 2002-12-24
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 5245
; TYPE: PRT
; ORGANISM: Streptomyces fradiae
US-10-329-079-11

Query Match 1.9%; Score 120; DB 12; Length 5245;
Best Local Similarity 20.2%; Pred. No. 1.8;
Matches 262; Conservative 138; Mismatches 434; Indels 464; Gaps 68;

QY 102 PSSAAPNLTRACNAAREPGRSGCQGP-----VDGAVTTGAEICTRL-----GL 147
Db 464 PAABPADEGLEAVCDTFARQAATPEAPAVGVPVALTFAEADARVSRRLILSRGA 523
QY 148 EPENTILVLTALFEKAVFMCNVPFLHYGLDIIHNGDVIRIPFPVQLFMDVNLV 207
Db 524 GPE---VRVAVCLDRLNMPITV-----LAVLSGAVHVP----- 556
QY 208 PPFNTHRSIGEGFVYPTFPYNTGLCHLIHDCVIAPMAVALRVNVTAVARGAHLAF- 266
Db 557 -DPRSPHER-----LAAVERD--VAPILLVAERATAAVALAAPVILV 597
QY 267 -DENHEGAV-----LPPDITYVFSQSSSG-----TTTARG-----A 297
Db 598 DDPSTEALIDALDPOVTDADRTAPLLPCHAAVYVHTSGSTRGPKGVTVDRHGRLLQA 657
QY 298 RENDVNSTKSPSG-GFERLASI-----MAADTALHA----- 330
Db 658 HRRVTFRIRDSAGPGRAAHVSSFPDASDPLLANVAGHELAWIDELRFPDPGVAY 717
QY 331 -----EVIPTNGYEE--TPT-----DIKEWPMFIGEGTLPRLNAL 365
Db 718 FDRRIDYVDLTPYFRSLDAGLLEEGFPFCSVALGGEAMDGELWERLRAAAPRV TAM 777
QY 366 GSY-----TARVA-----GVIGAMVFPNSALVLTVEDESGMTEAKDGGPSPFN 411
Db 778 NYTGTETAVDAVTVLGDLPFGTIGRPV--PRWAY--VLDAGLPVPGVLGELY-- 830
QY 412 FYQFAGPHLANPQDTRDGHVLSQSTGSSNTPESVDYALICGFGAP-----LLAR 463
Db 831 ---LAGPVARGYLQ---HALTAER-----FVACPFGKPGRMVTRTGLAR 871
QY 464 L-----LFLVLRCD-----AGAF-----TGGHGDALKYV 487
Db 872 WLPDGLHYVVGDEQVKIRGFRIEPOGEVAALRELGVAATAVTRDTPGTRRLVGYV 931

Fri Jan 30 16:10:01 2004

862 TIAAVKFTTEYAAI--NFINLPPTCIGELAQFYMANLILKYCDHSQYLINTLTSITGA 919
12499 LLATGTSCLAEVCVIGRSLEPTPT-----NMSAFISHTLSA---- 12536
920 RPRDPSSVLHWIRKDVTSAAADIETQAKALLEKTENLPELWTTAFTSTHVLVRAAMQRP 979
12537 -----PSDQV--VSRELQMASNVSAQORSYWRVQVFFRPPGRLARPHHPAYPCHELSQ 12589
980 VV-----LGISISKYHGAAGNNRVFOAGNWSGLNGKNCVPLFTFDSTRFIIACPRG 1032
12590 CVQDAWPSAVGCCSVRR---SVRSTMSLMVGSFQKMLAGTS-----RPRRF----- 12631
1033 GFICPVTGPSSGNRETTLSQDVRGIIIVSGAMVQLAIYATVVRVAVGARACHMAFDWL 1092
12632 -----SSRRRTTIVITESPCSTSG-----RRTCAVAGASRISCAI---CSR 12670
1093 TDDEFLARDLELHDQIIQTLETPWVEGALVAVKILDEKTTAGDGETPTNLAFNFDSC 1152
12671 TSSS-----TRRSSGD-----R 12684
1153 PSHDTTSNVLNIGSNISGS 1172
12685 PANDTTSASASPPGSSVSGA 12704

Search completed: January 30, 2004, 13:20:11
Job time : 49 secs

1903590KP1



ion date: 03-03-2004
 king Officer: HABDULEKADER - HUDA ABDULEKADER
 n: OIPEScanning
 sier: 10190359

al Date: 09-16-2002

o.	Doccode	Number of pages
	NPL	12 /
	NPL	4 /
	NPL	5 /
	NPL	8 /
	NPL	9 /
	NPL	4 /
	NPL	5 /
	NPL	2 /
	NPL	13 /
	NPL	6 /
	NPL	14 /
	NPL	6 /
	NPL	8 /
	NPL	40 /
	NPL	6 /
	NPL	7 /
	NPL	8 /
	NPL	4 /
	NPL	3 /
	NPL	4 /
	NPL	30 /
	NPL	8 /
	NPL	4 /
	NPL	8 /
	NPL	17 /
	NPL	14 /
	NPL	5 /
	NPL	9 /
	NPL	12 /
	NPL	7 /
	NPL	10 /
	NPL	
	NPL	
	NPL	
	NPL	